Analysis on the Metabolic Capabilities of five *Salmonella* Strains through Genome-Scale Metabolic Models

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Tong Ding

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Adviser: Dr. David Baumler

July 2017

Acknowledgements

I would like to thank my advisor Dr. David Baumler for providing me such a valuable opportunity to join his research team and pursue a Master's degree in Food Science. He is an ideal mentor and talented professor who knows how to organize a team, motivate the students, and create memorable fun moments. I'm always impressed that he can simultaneously play multiple roles including a good father, a bandleader, the knowledgeable Dr. Pepper, and a creative professor who is full of surprises.

I appreciate that Dr. Joellen Feirtag and Dr. Anup Kollanoor Johny would like to be my committee members and provide valuable suggestions.

I would also like to thank all the team Baumler members, as it was awesome to work with a group of intelligent, open-minded, and hard-working students. I would be even more thankful if they keep my fishes alive after I leave.

Thanks to the Department of Food Science and Nutrition, the College of Food, Agricultural, and Natural Resource Sciences, and the University of Minnesota, I was able to live, study, and think about life in a utopian environment away from unnecessary troubles.

Finally, I want to thank my parents who value and support my education, dear friends, dear faculty members, and all the kind-hearted people I have met during my study at the University of Minnesota.

General Introduction

In every country of the world, foodborne diseases caused by *Salmonella* represent a severe problem to the food supply as well as the public health. The work presented here in this dissertation, looks to investigate food safety related to sustainable farming practices, genome evolution of pathogenic bacteria during host-interactions, and harness post-genomic data to use systems biology methods to elucidate differentiating metabolic capabilities and targets of control of numerous *Salmonella* serovars.

The first chapter introduces detailed information about the background information about *Salmonella* as a foodborne pathogen. The second examines computational methods to determine if we can accurately predict genome evolution of pathogenic *Escherichia coli* and *Salmonella* during host interactions in niches in humans. The third chapter examines the food safety risks associated with the use of chicken manure for agricultural sustainable farming practices in Minnesota. Pathogenic bacteria including *Salmonella* are also a concern for sustainable farming in which organic fertilizers such as animal wastes are utilized. An analysis on microbiological hazards for such a sustainable farming system was presented in the third chapter. Finally, systems biology approaches were used in the study described in Chapter 4 to analyze strain to strain differences of metabolism of these pathogenic microorganisms.

Throughout evolution bacteria have gained or lost certain metabolic properties to better compete with other microorganisms in the changing living condition found in environmental niches found in hosts. Therefore, to develop advanced strategies fighting

against pathogenic bacteria, a solid understanding must be obtained on their capability to metabolize available nutrients within different hosts or environmental niches during infection. The genome-scale metabolic models (GEMs) constructed *in silico* allow us to conduct simulations mimicking real-life situation by interpreting complex bacterial metabolic systems to conduct predictions during bacteria-host/environment interactions. A publication reprinted in Chapter 2 presents work that we conducted to analyze the metabolism-related genes essential to various *Salmonella* and *Escherichia coli* species under simulated environments found in three niches where they cause disease.

Chapter 4 discussed a study on analyzing five different *Salmonella* strains' metabolic capabilities through a systems biology approach. The objective of the study was to gain a better understanding of differentiating metabolic capabilities among various *Salmonella* strains through efficient model construction and accurate prediction. Overall, the GEMs generated in this study can make good predictions when compared to experimental results, showing their great potentials in analyzing pathogenic bacteria and developing related pathogen control strategies, and the usefulness of this approach for the future examination of 100's to 1,000s of genomes of *Salmonella* spp..

Table of Contents

List of Tables

List of Figures

Chapter I A review of the literature on *Salmonella* and genome-scale metabolic model

I. 1 General Aspects of *Salmonella*

I. 1. 1 History and Prevalence

The organism *Salmonella* was named after Dr. Daniel Salmon who isolated the pathogenic bacterium from the intestinal tract of swine in 1885, and the microbial pathogen has since then been recognized as a challenge for public health. *Salmonella* can cause several types of infections, including mild and self-limiting gastroenteritis and typhoid fever caused by typhoid *Salmonella* that leads to serious symptoms. Among the ten deadliest outbreaks in the history of United States, *Salmonella* caused three and two of those outbreaks are linked to typhoid fever. The outbreak occurred in 1903 sickened 1,350 people leading to 82 deaths due to public water source contamination, and the pathogen again sickened more than 1,500 with 150 deaths linked to contaminated oysters in 1924 [1]. As the cause of typhoid fever in the early days with a mortality rate up to 1 in 10 before antibiotics were discovered, *Salmonella* raised public fear and distorted the public awareness on disease carrier.

Investigations on “Typhoid Mary” were carried out in 1900 when victims were found living in the same neighborhoods equipped with safe drinking water system, and all the evidence pointed to an asymptomatic *Salmonella* carrier who infected 51 people and caused 3 deaths. As a pathogen carrier displaying no infection symptoms, Mary Mallon worked as a cook and unconsciously spread the pathogen to her employers. The public health authorities forcibly kept and isolated her twice, and Mallon ended up living in quarantine at a hospital for the rest of her life [2]. This became one of the very first

cases that made the public reconsider the ethical conflict between individual liberty (Typhoid Mary) and the greater good of the society (public health), as it projected a prejudiced public attitude towards disease carriers in the early days when the knowledge of foodborne pathogens was not well elucidated.

Today, *Salmonella* is still one of the most dangerous foodborne pathogens causing massive outbreaks and numerous illnesses associated with hospitalization and even death. Although the incidence of typhoid fever has decreased dramatically, more and more non-typhoidal *Salmonella* infections occur (Figure 1) [3]. Four out of the ten biggest food recalls in the U.S. history were caused by *Salmonella*, including the Peanut Corporation of America (PCA)'s recall on 2,100 products that involve more than 200 companies in 2009. By the time the company was forced to make a recall, their products had already led to more than 714 illnesses and nine deaths in 46 states [1]. Along with these outbreaks, there were 62 related lawsuits from 1998 to 2014 including the 2009 PCA *Salmonella* outbreak [4].

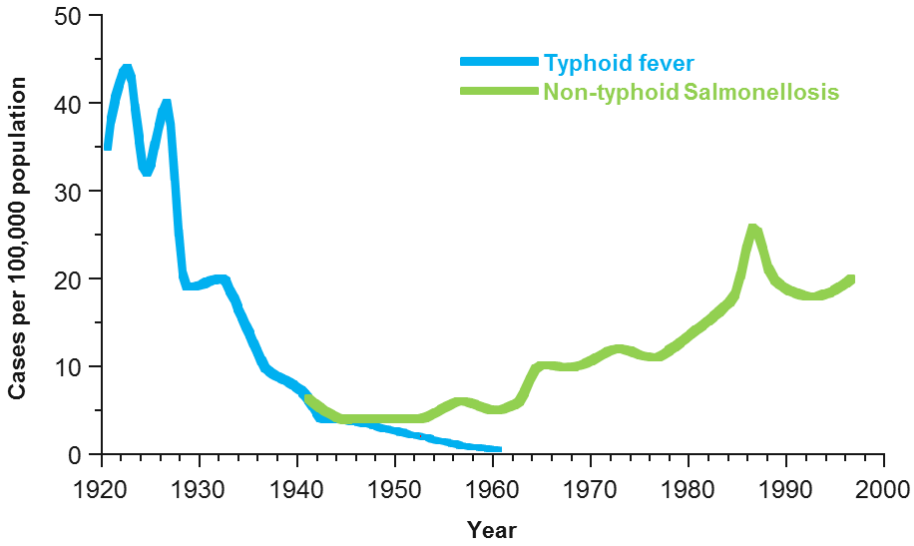

Figure 1: The incidence of typhoid fever and non-typhoid Salmonellosis (1920 – 2000)

According to the Centers for Disease Control and Prevention (CDC), an estimate of one million people get sick each year due to *Salmonella* in the U.S., resulting in approximately 19,000 hospitalizations and 380 deaths. Food related *Salmonella* strains are mainly non-typhoidal, such as the strain belonging to the non-typhoidal *Salmonella* serovar Typhimurium occurred in the PCA outbreak. This type of *Salmonella* was ranked the number one human food poisoning risk among all the bacterial pathogens tracked by the CDC from 2000 to 2008 (Table 1).

Table 1: The Centers for Disease Control and Prevention estimated annual illnesses, hospitalizations, and deaths due to bacterial pathogens in the United States (2000 – 2008)

Pathogen	Estimated annual illnesses	Estimated annual hospitalizations	Estimated annual deaths
Salmonella spp., nontyphoidal*	1,000,000	19,000	380
Clostridium perfringens, foodborne	970,000	440	26
Campylobacter spp.	850,000	8,500	76
Streptococcus spp. *group A*, foodborne	240,000	1,100	6
Shigella spp.	130,000	1,500	10
E. coli (STEC) non–O157	110,000	270	1
Yersinia enterocolitica	98,000	530	29
Bacillus cereus, foodborne	63,000	20	0
E. coli (STEC) O157	63,000	2,100	20
V. parahaemolyticus	35,000	100	4
Enterotoxigenic *E. coli (*ETEC)	18,000	12	0
Vibrio spp., other	18,000	83	8
Diarrheagenic *E. coli* other than STEC and ETEC	12,000	8	0
Streptococcus	11,000	1	0
S. enterica serotype *Typhi*	1,800	200	0
Listeria monocytogenes	1,600	1,500	250
Brucella spp.	840	55	1
V. vulnificus	96	93	36
Vibrio cholerae, toxigenic	84	2	0
Mycobacterium bovis	60	31	3
Clostridium botulinum, foodborne	55	42	9

* Food related *Salmonella* strains are mainly non-typhoidal

The prevalence data on *Salmonella* can be found on the Foodborne Diseases Active Surveillance Network (FoodNet), which was established by the US Department of Agriculture, the Food and Drug Administration, the CDC, and 10 state health departments together to monitor foodborne pathogen outbreaks [5]. Among the nine food-borne pathogens tracked by FoodNet, *Salmonella* is responsible for up to half of the outbreak hospitalizations and deaths, costing an estimated $365 million in direct healthcare per year [6]. Compared to other foodborne pathogens such as *Escherichia coli* O157, there was no significant reduction of foodborne illnesses caused by *Salmonella* over the last decade. The incidence of foodborne *Salmonella* disease outbreak has increased in recent years (Figure 2). Even though the better rapid detection methods have been developed, the increase in consumer awareness on foodborne *Salmonella* disease leads to increased reporting to doctors and state or national public health departments, and the improvement in efficient microbial tracking system can explain the increase in overall recorded pathogen outbreaks, *Salmonella* outbreaks are still increasing.

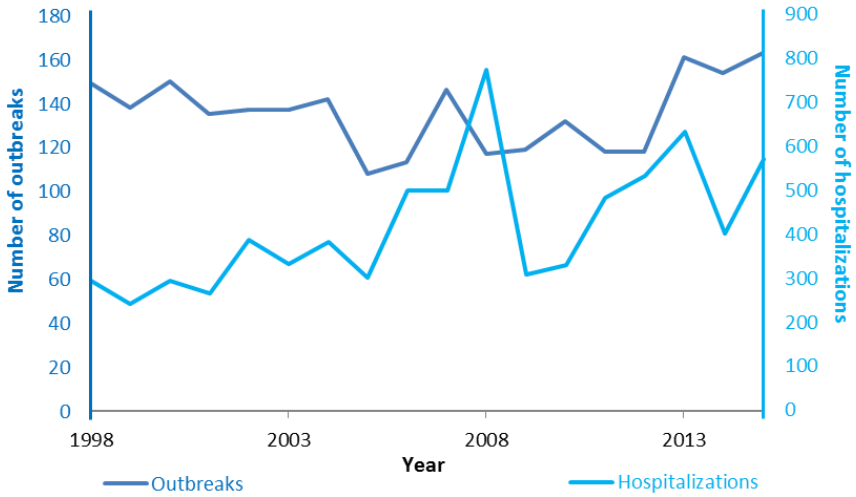

Figure 2: The number of outbreaks and hospitalizations caused by *Salmonella* in the United States from 1998 to 2015 recorded by the CDC

Furthermore, *Salmonella* is also a dangerous food security risk given it has caused the largest bioterrorist attack in the U. S. history. In the biological attack occurred in 1984, more than 751 people were infected with *Salmonella* serovar Typhimurium after eating or working at 10 local restaurants in The Dalles, Oregon. By tracing back the source of the *Salmonella* strain responsible for the outbreak to a religious commune, a bioterrorist plot was disclosed that intended to incapacitate the voting population in an attempt to gain political control of Wasco County. It took months for the public health authorities to investigate the cause, and fortunately no one was killed in the attack [7, 8].

With the well-established public health surveillance system and advanced detection technologies today, we are able to figure out the cause and trace back the source for pathogen contamination much more rapidly. However, the food system is facing many new food safety challenges as new emerging foodborne diseases are increasing and the food supply chain becomes more and more complex and global. To minimize production costs for foods, item can be manufactured in foreign countries or be made with imported ingredients as a result of globalization, which makes it harder to identify potential microbial hazards throughout the complicated food supply chain. In addition, products now can be transported nationwide or worldwide to serve more consumers, and that further hampers the traceability of the source for pathogen contamination. Hence, more efficient pathogen control strategies need to be developed to ensure food quality and to protect the health of consumers.

I. 1. 2 Characteristics and Taxonomy

Salmonella spp. are Gram-negative rod-shaped facultative anaerobes belonging to the *Enterobacteriaceae* family. These bacteria are non-spore forming, and most of them are motile with flagella. Most *Salmonella* strains can grow between 7 – 48 ºC, and several studies report that some strains can grow at temperatures as low as 4 ºC. Thus, this pathogen is able to survive in frozen conditions for extended periods such as in processed poultry products [9]. *Salmonella* can be acid tolerant, surviving pH ranges from 3.7 to 9.5. This pathogen can also survive A_w down to 0.94 and even on dry food like chocolate that has high fat content but is low in moisture [10]. Moreover, *Salmonella* can tolerate up to 8% salt in foods. The severity of *Salmonella* infection varies by the infectious dose, the host health status, and different serotypes' pathogenic capabilities. The typhpidal *Salmonella* Typhi and Paratyphi A can cause typhoid fever with fewer than 1,000 cells, while the non-typhoidal *Salmonella* species can cause infection with as low as one cell [11]. Unlike typhoidal *Salmonella* that is problematic mostly in developing countries, non-typhoidal *Salmonella* is more of a global concern as it leads to about 93.8 million illnesses and 155,000 deaths annually throughout the world. In the U.S., the outbreaks caused by typhoid *Salmonella* has been decreasing since 1996, while non-typhoidal *Salmonella* remains a concern.

A previous study found that 87% of *Salmonella* outbreak cases were foodborne, compared to person-to-person infection (10%) and infection caused by pets (3%) including dogs, cats, reptiles, amphibians, guinea pigs, hamsters, and birds [12]. The typhoidal *Salmonella* can be transferred from water or food items to human, and humans are its only host identified so far. Non-typhoidal *Salmonella* strains on the contrary have

a broad animal host range that includes a variety of farm animals like chickens, cattle, and pigs, and therefore animal products such as meat, poultry, dairy, and eggs are most susceptible to *Salmonella* contamination. According to the CDC, most outbreaks caused by *Salmonella* are linked to poultry (29%), eggs (18%), pork (12%), and beef (8%). This pathogen can also be transferred from animal carriers to the environment through fecal matter which may lead to contamination of produce via irrigation water or soil splash back during periods of heavy rain. Consequently, there is a large percentage (>13%) of *Salmonella* outbreaks relates to vine vegetables, fruits, nuts, leafy greens, roots, and sprouts (Figure 3).

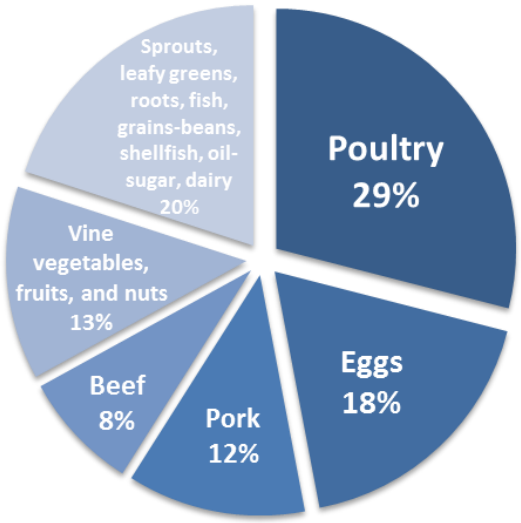

Figure 3: Food items associated with *Salmonella* disease outbreaks according to the CDC National Outbreak Reporting System (2004–2008)

The symptoms of *Salmonella* infection range from typhoid fever caused by typhoidal *Salmonella* strains to gastroenteritis caused by nontyphoidal *Salmonella* strains, and most human non-typhoidal *Salmonella* infections were caused by *Salmonella* serovar Typhimurium or serovar Enteritidis acquired from contaminated food sources [13].

Typical symptoms of non-typhoidal *Salmonella* infection, or Salmonellosis, include diarrhea, fever, and abdominal cramps that may appear 6 to 72 hours after ingestion of contaminated food or water and the symptoms may last 2 to 7 days. Although most Salmonellosis infections are self-limiting, meaning most people would recover without any treatment, some diarrhea cases may lead to serious dehydration that requires hospitalization and electrolyte replacement. Healthy adults may be asymptomatic carriers, while the high-risk population may suffer from prolonged complications such as reactive arthritis. The high-risk population for *Salmonella* infection includes infants fed with formula instead of breast milk, children under 5, adults above 65, pregnant women, people taking medications, and the immunocompromised individuals. The population demographics from FoodNet indicate that children are more at risk for foodborne pathogen infection, while the elderly (> 70 yrs) are more vulnerable to this pathogen to cause illness [5].

About 1200 *Salmonella* strains were identified by 1967, and the number doubled after 50 years. The current nomenclature system categorizes *Salmonella* (*S.*) into two species (spp.), *S. bongori* and *S. enterica*, and the later can be further divided into six subspecies (I – VI), namely *enterica, salamae, arizonae, diarizonae, houtenae,* and *indica* [14]. In addition to the serotypes that were classified upon biochemical characterization, there are more than 2,500 serovars in the *Salmonella* genus that are differentiated by their antigenic properties (Table 2). Although most serovars belonging to the subsp. enterica are widely known by names related to the geographic area where they were isolated, the Kauffman–White serotyping system was applied to form the basis

of a global understanding on *Salmonella* nomenclature for all *Salmonella* surveillance networks. By serotyping *Salmonella* spp. based on two surface structures, somatic (O) and flagellar (H) antigens, this taxonomy system has been widely used by laboratories as a gold standard for the characterization of *Salmonella* isolates. This serotyping scheme can recognize up to 2,523 serotypes based on the combinations between 46 O and 114 H antigenic serogroups [15].

Table 2: Current *Salmonella* nomenclature system based on the Kauffman–White serotyping scheme

Genus	Species	Subspecies	Serovar example(s)	Number of serotypes	Sources
Salmonella	*enterica*	I (*enterica*)	Paratyphi, Typhi, Enteritidis, Typhimurium	1504	Warm-blooded animals
		II (*salamae*)	9,46:z:z39	502	Cold-blooded animals and the environment
		IIIa (*arizonae*)	43:z29:-	95	Cold-blooded animals and the environment
		IIIb (*diarizonae*)	6,7:l,v:1,5,7	333	Cold-blooded animals and the environment
		IV (*houtenae*)	21:m,t:-	72	Cold-blooded animals and the environment
		VI (*indica*)	59:z36:-	13	Cold-blooded animals and the environment
	bongori	V	13,22:z39:-	22	Cold-blooded animals and the environment
	subterranea				
				Total: 2541	

The Kauffman–White classification approach is applicable to specify taxonomic differentiation and pathogenic grouping across the wide range of *Salmonella* spp.;

however, incomplete expression of antigens or novel genetic combinations will decrease the accuracy or limit the use of this system [16]. For instance, a newly found strain that expresses only a partial antigenic type cannot be assigned with a definitive serovar name based on this nomenclature scheme, such that we may fail to identify new *Salmonella* strains by misinterpreting the strain's phenotypic information. Nowadays, the robust development of genomic techniques including whole genome sequencing (WGS) and multilocus sequence typing (MLST) methodologies have been considered replacements for the traditional serotyping approach for *Salmonella* [17, 18]. In this review, the naming of *Salmonella* strains follows the traditional nomenclature and the abbreviated names will be used by simply listing the serotype as in most of the literature.

I. 1. 3 Control and Treatment

Theoretically, diseases and illnesses caused by foodborne pathogens are largely preventable. The preventive measure against *Salmonella* recommended by the WHO is to conduct good food hygiene practices during food handling and cooking. However, many factors can be easily ignored throughout the food supply chain that results in *Salmonella* contamination. In 1994, up to 224,000 U.S. citizens in more than one state developed syndromes of gastroenteritis after they consumed Schwan's ice cream. Investigations found the ice cream premix was contaminated during transportation, where the tanker trailers had previously carried unpasteurized liquid eggs containing *Salmonella* Enteritidis [19]. Six years later, a new FDA rule was issued to prevent *S.* Enteritidis contamination on shell eggs, for which producers are responsible for pathogen control from production to transport and storage. Even so this pathogen remains the leading

cause of foodborne illnesses linked to poultry products in the U.S., and it once led to a nationwide recall involving 500 million eggs after another multi state outbreak in 2010. Control of *Salmonella* infection is usually difficult due to the pathogen's tolerance and adaptation to environmental stress.

The clinical treatments for Salmonellosis present to date are limited to fluid/electrolyte therapy and antibiotics. The former is generally used to treat dehydration as a result of severe diarrhea, while the latter has been extensively used for complications such as bacteremia caused by the pathogen invading the host's bloodstream. *Salmonella* strains have already developed resistant capabilities against numerous antimicrobials due to the overuse of antibiotics, and the emergence of these drug-resistant strains poses a huge public health concern. There are now an estimated 50,000 deaths caused by antibiotic-resistant pathogenic bacteria each year in the Europe and the U.S., while hundreds of thousands more occur worldwide [20].

Our control over pathogen infection is diminishing as the rate of new drug discovery slows while the incidence of antibiotic-resistant infections rises. Systematically screening a wide variety of compounds used to be a common strategy for discovering antimicrobial agents for centuries. Up to 80% of the world's population relies on plant-derived medicines considering their minimized side effects, and researchers have found antibacterial properties in more than 64 plants [21]. However, the discovery of new antibiotics only predominated from the 1960s to the 1990s, peaked in 2000, and then declined sharply thereafter. Up to 2014, about one-third of the discovered antibiotics became obsolescent due to the development of drug resistance in pathogens [22]. The

widespread and overuse of antibiotics led to selective evolutionary pressure on pathogens which resulted in the acquisition of related drug-resistance genes to survive in the presence of antibiotics.

Moreover, the use of antibiotics may lead to dysbiosis in the human microbiome. A healthy balanced gut microbiome plays vital roles in human metabolic and immune systems. Once the ecology is disrupted, the weakened nutrient absorption and microbial defense may result in health problems due to metabolic, immunological, or developmental disorders [23]. A murine study on *Salmonella* Typhimurium infection showed the use of antibiotics can profoundly alter the microbiota composition and increase the host's susceptibility to pathogen intestinal colonization, predisposing the host to more severe infection [24]. The unintentional consequences of antibiotic over usage indicates we are reaching the end of the antibiotic era.

Researchers today are trying to discover compounds that target molecular targets to mute or eliminate disease-causing bacteria with the help of advanced genomic techniques [25]. Along with these techniques, genetic data has been widely used for identifying outbreak pathogen, tracking contamination sources, analyzing conservation of virulence factors, and predicting antimicrobial resistance. WGS has been recently adopted as a surveillance tool by the National Antimicrobial Resistance Monitoring System (NARMS), collaboration among state and local public health departments such as the CDC, the FDA, and the USDA. This national public health surveillance system started to record the genetic data for *Salmonella* isolated from patients in 2014 to identify drug-resistant genes and mutations. NARMS reported that 4.3% of the non-typhoidal

Salmonella isolates they collected had decreased susceptibility to ciprofloxacin, among which *S.* Enteritidis was the most common serotype. 2.4% of the isolates were resistant to ceftriaxone, including the isolates belonging to *S.* Typhimurium and *S.* I 4,[5],12:i:-. Fluoroquinolones such as ciprofloxacin and third-generation cephalosporins including ceftriaxone are first-line treatment agents to treat *Salmonella* infection, nevertheless an increasing percentage of *Salmonella* isolates with decreased susceptibility to these antibiotics has been observed over the last decade. NARMS also reported that 3.1% of those non-typhoidal isolates developed multidrug-resistance to at least ACSSuT (ampicillin, chloramphenicol, streptomycin, sulfonamide, and tetracycline), including the isolates belonging to *S.* Typhimurium. Particularly, up to 42.7% of the *S.* I 4,[5],12:i:- isolates tested by NARMS were resistant to ASSuT but not chloramphenicol [26]. Genomic techniques including WGS have proven useful in pathogen surveillance and food safety protection, and the long term goal for this project is update current *Salmonella* treatment methods for a more promising clinical outcome through genomic approaches.

I. 1. 4 Representative Strains

All the strains used in this study belong to *Salmonella* spp. *enterica* subsp. I (*enterica*), the major cause of human and animal *Salmonella* infections (Table 3). Three out of the five strains are under the non-typhoidal serovar Typhimurium and Enteritidis which most human inflammatory gastroenteritis or so-called food poisoning is linked to [27]. The strain P125109 was first isolated from poultry [28]. The serovar it belongs to, Enteritidis, is highly associated with contaminated eggs and poultry products because this

serovar may have adapted to the reproductive organ of laying hens [29]. Both LT 2 and UK 1 strains are within the serovar Typhimurium, which causes symptoms similar to Typhi in the murine model. This serovar is globally distributed and it can cause disease in a variety of hosts. The *Salmonella* strain Universal Killer (UK) 1 can cause lethal infection among a variety of animals such as chickens, mice, calves, pigs, and horses. This strain was first isolated from a chick that was inoculated with a strain found on a deadly infected horse in 1991 [30]. The wild-type strain LT 2 was named based on Lilleengen type (LT) described in an early study in 1960, and this strain was found in a variety of hosts including human [31, 32].

The monophasic serovar 4,5,12:i:- initially isolated from swine is a variant of *Salmonella* Typhimurium as it lacks the second flagellar antigen. This serovar has emerged since the 1990's and it became a common serovar linked to swine products in the past decades [33, 34]. Most importantly, this pathogenic strain showed multidrug resistance to antibiotics including ampicillin, streptomycin, sulfisoxazole, and tetracycline. Antibiotic resistance in pathogen may lead to treatment failure and development of complications such as bloodstream infection in patients, and therefore *S.* 4,5,12:i:- represents a current public health risk.

Salmonella serovar Abaetetuba was first isolated in creek waters of the Zaimán arroyo in Argentina, and this pathogen is able to cause infections in reptiles such as marine lizards. [35, 36]. It was considered one in six reptile-associated serovars of the top 20 emerging serovars from 1987 to 1997 [37]. The increase of *Salmonella* human infections caused by rare serovars including *Salmonella* Abaetetuba speculated the

potential emergence of *Salmonella* from the environment and exotic pets [38, 39]. In England and Wales, 58 salmonellosis cases were reported from 1981 to 2013 that were associated with *S.* Abaetetuba. It was found in raw and ready-to-eat bean sprouts and sprouted seeds on retail sale in England and Northern Ireland in 2011 [40]. Nevertheless, this serovar has not caused any outbreaks in the U.S to date, which makes it a good reference serovar in this study.

Table 3: Isolation sources and related outbreaks of the five *Salmonella* strains in interest

Strain	**Serovar**	**Isolation sources**	**Outbreaks in U.S.**
LT2	Typhimurium	Human, horse, cattle, guinea pig, rat and mouse, 1960	Frozen Feeder Rodents (2014) Live Poultry (2013) Hedgehogs (2012) Cantaloupe (2012) Ground Beef (2011, 2013) African Dwarf Frogs (2011) Peanut Butter (2009) Tomatoes (2006)
UK1	Typhimurium	Horse, chicken, 1991	
CVM23701	4,5,12:i:-	Swine products (pork), 2003	Pork (2015) Alfalfa Sprouts (2008) Frozen Rodents (2008) Banquet Pot Pies (2007)
P125109	Enteritidis	Poultry products (chicken meat, liver, and egg), 2012	Live Poultry (2015) Frozen Chicken Entrees (2015) Bean Sprouts (2014) Ground Beef (2012) Turkish Pine Nuts (2011) Alfalfa and Spicy Sprouts (2011) Shell Eggs (2010)
ATCC 35640	Abaetetuba	Creek water, 1983	None

I. 2 **Systems Biology and *Salmonella***

I. 2. 1 Systems Biology on a Genome Scale

Genomic data have been greatly applied for systematic study on microorganisms. The genome of bacterial chromosome encodes their internal metabolic characteristics, and therefore by studying their genome information we can depict a blueprint describing the potential traits the organisms possess that help them survive under different environmental pressures. With a better understanding of their metabolic capabilities, predictions can be made to disclose bacterial behaviors. Accordingly, it is possible to computationally reconstruct metabolic networks using the genotype information for microorganisms of interest and predict their phenotypic behaviors under simulated environment computationally. The genome-scale reconstruction (GENRE) of a metabolic network is enabled with the increase in the genomic data through whole-genome sequencing along with the genotype-phenotype studies.

Most importantly, we can discover detailed genetic relationships between different strains and reveal ancestral lineage within species through genome-scale analysis, such as the high-throughput DNA sequencing technique that has been applied in modern public health related microbiology study to extend and partially replace the traditional nomenclature because it provides more accurate phylogenetic relationships between isolates [16, 17]. Genome-scale analysis is also used in virulence or antimicrobial resistance gene identification. The bacterial genome changes continuously evolutionarily over time, and by determining the genome level changes one can better understand the differentiating metabolic and virulence factor capacity for an organism.

The genome data of many microorganisms can be retrieved from online databases that are open to the public. The availability of genome sequencing data of microorganisms, including pathogens such as *Salmonella*, provides various opportunities to systematically or partially study these organisms through a systems biology approach. All the genome data used for the comparable alignment in this study were obtained from the National Center for Biotechnology Information (NCBI), considering it is a trustworthy publicly accessible online databank funded by the U.S. government for molecular biology information (https://www.ncbi.nlm.nih.gov/).

To visually compare the genomes of the five *Salmonella* strains used in this study, the genome alignment tool Mauve was chosen owing to its multiple strengths. Mauve has for long been used as an alignment system that integrates multiple genome sequence alignment with large-scale evolutionary analysis. Unlike other alignment methods, Mauve's anchoring algorithm identifies and aligns homologous region of sequence shared by two or more of the aligned genomes into locally collinear blocks (LCBs). Thus, Mauve is able to detect gene deletion, insertion, rearrangement or inversions in conserved regions across multiple genomes with high sensitivity and convenience. The Graphical User Interface (GUI) developed for Mauve is very user-friendly for displaying the rearrangement structure of multiple genome sequences [41].

I. 2. 2 Systems Biology with Metabolic Models

Systems biology uses computational and mathematical modeling to understand complex biological systems. Embedded with countless interactions among various cellular components, the complexity of biological systems is complicated even for a

simple tiny bacterial cell. To reveal the complete set of all genome-scale metabolic reactions along with the molecular compounds (metabolites) involved in cell activities, mathematical modeling has been developed since 1960s as an integrated approach based on physicochemical laws such as the biochemical systems theory and metabolic control analysis. Through computational modeling, bacterial genotype is converted into a stoichiometric matrix to contain all of the metabolic reactions and metabolites present into an *in silico* metabolic system, in which the metabolic capabilities can be identified by defining systemic constraints of the network and using linear and mixed integer programming and optimization methods and algorithms.

A model constructed this way is called a genome-scale metabolic model (GEM) in this study, and it can be treated as a functioning network consisting of metabolites and reactions annotated by the bacterial genome that are essential for a bacterial cell to live and generate biomass under certain environments. Computationally, the metabolic network is encoded as a stoichiometric matrix with a list of all compounds and metabolic reactions that are transported into and out of the cell. The biomass equation is a metabolic reaction that contains all the precise amounts of metabolites necessary to produce 1 g of dry cell weight. Optimization algorithms can predict the maximum biomass production from growth on each metabolite or complex mixtures that enable the simulation of cell growth and the products that contribute to the generation of biomass for cell growth. By setting organism-specific features such as O_2 uptake rate and environmental constraints, the GEM can be used to make predictions related to the differentiating capabilities of each strain's metabolism, and also to new targets for control and growth prevention.

GEMs have been proven useful in understanding the metabolic properties of a variety of organisms. The first GEM was constructed in 1994 for *Haemophilus influenzae*, and now GEMs have been built for a wide range of organisms. However, there are few GEMs built for foodborne pathogens, limiting to *Staphylococcus aureus*, *Vibrio vulnificus*, *Escherichia coli*, and a single strain of *Salmonella* Typhimurium [42-45]. *As* one of the best-studied bacteria, *E. coli* owns several GEMs that have been expended and developed throughout the past two decades. The GEMs for *E. coli* have been used as scaffolds to construct GEMs for other microorganisms through gene orthology and bioinformatics sequence similarity analysis, such as iMA945 for *Salmonella* Typhimurium strain LT2 [46]. Other than *i*MA945, iRR1083 was built for the same *Salmonella* strain LT2 by another research team in the same year [47, 48]. Later a team of more than 20 experts in systems biology collaborated to generate a comprehensive model (STM_v1.0) for *Salmonella* LT2 by reconciling these two models [43]. So far there is only one complete GEM built for one of the pathogenic *Salmonella* strains, and this study leads to the construction of 5 GEMs for different *Salmonella* strains to examine new post-genomic similarities and differences related to the metabolic capabilities.

The applications of GEMs include (1) metabolic engineering that leads to increased/decreased metabolite production, (2) directing biological discovery based on comparison of *in silico* predictions and experimental outcomes, (3) examination and quantitative interpretation of metabolic physiology, (4) analysis of network properties, and (5) revealing bacterial evolution [49]. Therefore, the ultimate goal for this study is to

comprehensively simulate cellular systems with dynamic models through a systems biology approach for these future applications. Specifically, the GEMs constructed in this study will be used to investigate metabolic features during host-pathogen interactions, to identify differentiating metabolic properties between different strains, and to identify new targets for control of this pathogen and possible new treatment methods of associated infections.

I. 2. 3 GEM Construction

As shown in Figure 4, the first step of GEM construction is to reconstruct a draft metabolic model based on the gene-protein-reaction association of an organism, which requires gene annotation. Genome annotation is to identify coding regions, open reading frames or features (used in KBase) to assign putative functions to genes with database, and these genes may be involved in central metabolism, carbohydrate assimilation, and energy generation. Once expressed in equations, the complete set of gene functions, genotype, allows the reconstruction of a theoretical metabolic network and the related stoichiometric matrix. With the fast development of sequencing techniques, the complete genome with a size of several million base pairs now can be sequenced in a short time for any organism, and most of the sequence data are available online. These sequenced genomes then can be annotated to reconstruct the metabolic network for an organism, largely contributing to the development of GEM. Besides, the continuous exploration of uncharacterized genes will help improve the genotype complicity, thus improving the accuracy of gene annotation and the quality of the related GEMs.

The U.S. Department of Energy Systems Biology Knowledgebase (KBase) was used to construct draft models in this study (https://kbase.us/). KBase is a semi-automatic modeling platform that allows users to conduct *in silico* experiment, such as GEM construction, with high reproducibility on its Narrative interface. Its model reconstruction pipeline uses re-annotated genome sequences to build draft models. Specifically, the Rapid Annotation using Subsystem Technology (RAST) annotation algorithm is applied to conduct structural and functional annotation for the genomes uploaded onto the platform. The RAST server annotates bacterial and archaeal genomes by identifying protein-encoding genes and assigning gene functions for later metabolic network reconstruction based on represented subsystems in the genome. Compared to other algorithms, RAST possesses higher accuracy, consistency, and completeness due to the continuously updating of the subsystems library [50].

The re-annotated genomes then can be processed with the Model SEED pipeline, a bioinformatics algorithm for metabolic model reconstruction associated with subsystems, to generate a preliminary model that consists of intracellular and extracellular metabolites as well as transport reactions according to the RAST annotation [51, 52]. After that an analysis-ready model is generated to simulate biomass production using only transportable nutrients and strain-specific biomass reaction. Later the model can be gapfilled on a specific growth condition to complete the reaction network based on Flux Balance Analysis (FBA) using an optimization-based approach without referring to the genome [53]. To conduct FBA, the example network shown in Figure 4 will be first converted into a stoichiometric matrix ($\boldsymbol{S}$), in which each row represents a metabolite,

each column represents a reaction (***v***), and the numbers are the stoichiometric coefficients for metabolites participating in reactions. At steady state, the mass balance equation can be set as $\boldsymbol{S} \bullet \boldsymbol{v} = 0$, assuming the amount of metabolites entering the network equals to the amount leaving. Along with other constraints that set the reaction bounds ($a_i < v_i < b_i$), an allowable solution space can be achieved mathematically. Then, one can determine a single optimal solution by optimizing an objective function through linear programming. For example, the objective function on biomass generation was maximized in this study.

The final curation of the model requires manual addition or removal of reactions based on the comparison of phenotype prediction using FBA and experimental confirmation with Biolog phenotype microarrays. Manual curation is required to obtain high-quality GEM with accurate metabolic predictive power, because the metabolic networks cannot be reconstructed completely automatically to make accurate predictions with the current knowledge on genome annotations [54].

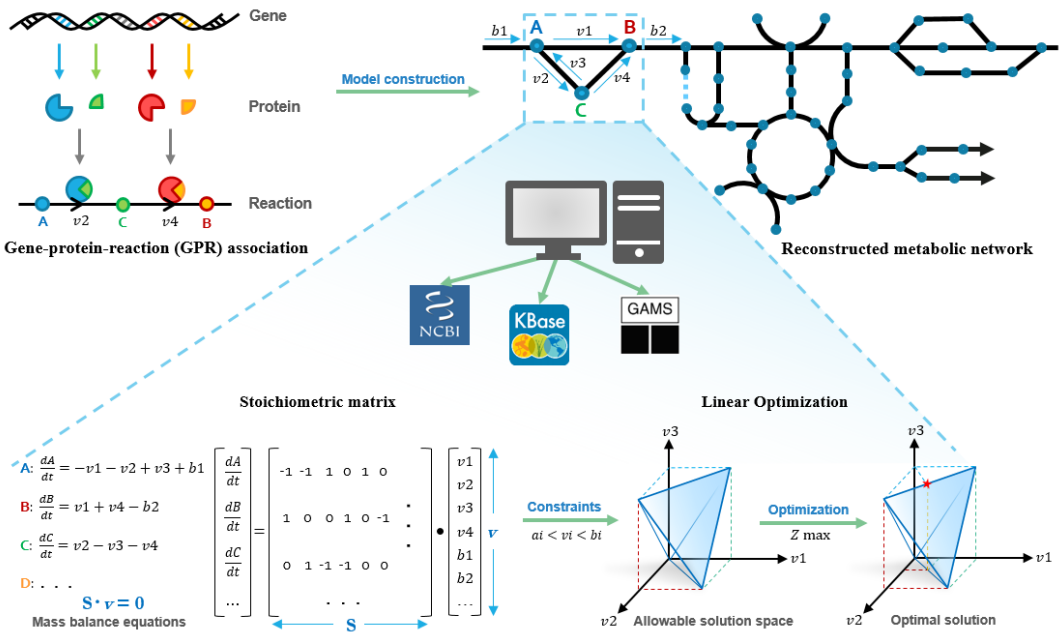

Figure 4: Flow diagram of the GEM construction method used in this study

I. 2. 4 GEM Prediction

Metabolic FBA was developed to simulate cellular activities in a metabolic network with physicochemical constraints. Although the information on cellular dynamics and metabolic regulations is limited, FBA allows accurate simulation based on the stoichiometry of metabolic pathways and the metabolic demands, and it can incorporate additional information when it is available. By contrasting the genotype-phenotype relationship of different *Salmonella* strains, we will be able to explore how these different strains have adapted to their preferable environmental or host niches. Hence, FBA is a powerful tool for model analysis.

FBA can also be used to predict reactions that are required to perform essential cellular activities allowing a microorganism to be viable and reproducible in a given environment. These reactions are referred to as essential reactions, and their corresponding genes are potential targets for research on growth control of microorganisms. A metabolic model of *Clostridium difficile* strain 630 was constructed by Larocque *at el.* to explore therapeutic targets against *Clostridium difficile* infections through gene essentiality simulation [55]. The identification and comparison of essential reactions and genes among different *Salmonella* strains under various conditions will provide new ideas on how to eliminate them in environments that requires food safety.

In addition, the metabolic network of a microorganism is complex, and sometimes the GEM may fail to make accurate predictions if it is built solely based on metabolic reactions. For instance, a bacterial genome possess a gene that is supposed to encode certain enzyme and contribute to a metabolic reaction while the reaction may not actually be present in that strain's metabolic system due to interference of repressors. Therefore,

the epistatic interactions between genes must be experimentally verified to construct a comprehensive metabolic network model. In this study, experimental data were obtained to qualitatively validate the GEMs by adding or removing metabolic reactions to eliminate false negative and false positive predictions. By combining the computational and experimental approaches, systems biology can be used to study the complex *Salmonella* metabolic networks systematically at a genome level.

Chapter II Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments

Publication reprint

Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments

Tong Ding[1☯], Kyle A. Case[1☯], Morrine A. Omolo[1☯], Holly A. Reiland[1☯], Zachary P. Metz[1☯], Xinyu Diao[1☯], David J. Baumler[1,2,3]*

1 Department of Food Science and Nutrition, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, 2 Microbial and Plant Genomics Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, 3 Biotechnology Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America

☯ These authors contributed equally to this work.
* dbaumler@umn.edu

Abstract

Microorganisms have evolved to occupy certain environmental niches, and the metabolic genes essential for growth in these locations are retained in the genomes. Many microorganisms inhabit niches located in the human body, sometimes causing disease, and may retain genes essential for growth in locations such as the bloodstream and urinary tract, or growth during intracellular invasion of the hosts' macrophage cells. Strains of Escherichia coli (E. coli) and Salmonella spp. are thought to have evolved over 100 million years from a common ancestor, and now cause disease in specific niches within humans. Here we have used a genome scale metabolic model representing the pangenome of E. coli which contains all metabolic reactions encoded by genes from 16 E. coli genomes, and have simulated environmental conditions found in the human bloodstream, urinary tract, and macrophage to determine essential metabolic genes needed for growth in each location. We compared the predicted essential genes for three E. coli strains and one Salmonella strain that cause disease in each host environment, and determined that essential gene retention could be accurately predicted using this approach. This project demonstrated that simulating human body environments such as the bloodstream can successfully lead to accurate computational predictions of essential/important genes.

PLOS | ONE

RESEARCH ARTICLE

Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments

Tong Ding[1☯], Kyle A. Case[1☯], Morrine A. Omolo[1☯], Holly A. Reiland[1☯], Zachary P. Metz[1☯], Xinyu Diao[1☯], David J. Baumler[1,2,3]*

1 Department of Food Science and Nutrition, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, **2** Microbial and Plant Genomics Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, **3** Biotechnology Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America

☯ These authors contributed equally to this work.
* dbaumler@umn.edu

OPEN ACCESS

Citation: Ding T, Case KA, Omolo MA, Reiland HA, Metz ZP, Diao X, et al. (2016) Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments. PLoS ONE 11(2): e0149423. doi:10.1371/journal.pone.0149423

Editor: Yousef Abu Kwaik, University of Louisville, UNITED STATES

Received: November 4, 2015

Accepted: January 31, 2016

Published: February 17, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Microorganisms have evolved to occupy certain environmental niches, and the metabolic genes essential for growth in these locations are retained in the genomes. Many microorganisms inhabit niches located in the human body, sometimes causing disease, and may retain genes essential for growth in locations such as the bloodstream and urinary tract, or growth during intracellular invasion of the hosts' macrophage cells. Strains of *Escherichia coli* (*E. coli*) and *Salmonella* spp. are thought to have evolved over 100 million years from a common ancestor, and now cause disease in specific niches within humans. Here we have used a genome scale metabolic model representing the pangenome of *E. coli* which contains all metabolic reactions encoded by genes from 16 *E. coli* genomes, and have simulated environmental conditions found in the human bloodstream, urinary tract, and macrophage to determine essential metabolic genes needed for growth in each location. We compared the predicted essential genes for three *E. coli* strains and one *Salmonella* strain that cause disease in each host environment, and determined that essential gene retention could be accurately predicted using this approach. This project demonstrated that simulating human body environments such as the bloodstream can successfully lead to accurate computational predictions of essential/important genes.

Introduction

Computational modeling has been widely used as an efficient approach in microbiology, which introduces mathematical components including variables, parameters, and equations in network constructions to reflect the behavior of organisms. Numerous types of networks have been constructed including signaling, regulatory, and metabolic pathways for organisms

ranging from microorganisms, such as *E. coli*, to multi-cellular eukaryotic organisms. By constructing genome-scale metabolic models (GEMs), the nature of an organism can be explored through computational analysis of its genome content. The *E. coli* K-12 strain MG1655 has had extensive computational metabolic networks generated for it so far, and its existing models are quite advanced that contain >2,000 reactions, >1,000 genes, and >1,000 metabolites [1–6]. These genome-scale models have been used for many studies that have guided the engineering of strains for increasing valuable end-products, promoting enzyme discovery, providing insight into the genome evolution of other enterobacteria [7,8], and leading to a new understanding of the connectivity, or coupling, of all the metabolic reactions and corresponding genes within the cell.

Currently, numerous *E. coli* metabolic networks have been constructed for commensal, enterohemorrhagic, and extra intestinal pathogenic strains [1,4]. Unlike studies using *E. coli* metabolic models, a *Salmonella* Typhimurium LT2 metabolic model was used to examine metabolic reactions and the corresponding essential genes that are necessary for cell viability during the infection process under simulated conditions inside the host [9]. The evolutionary process that leads to genome changes is based on the theory of natural selection, which states that in a given environmental niche, there is constant pressure to retain genes that are important for growth and survival in that particular condition. When the availability of nutrients in a host-cell environment can be used to further define the mathematical constraints for the metabolic model mimicking host-cell nutrient environment, a technique termed flux balance analysis (FBA) was used that identified 417 reactions used by *S. typhimurium* LT2 during human infection [9].

To systematically explore genes predicted as essential and important for cell growth in a given environment, we used an approach that focused on three main components: 1) generating a metabolic network and corresponding metabolic model representing the metabolic capabilities of the *E. coli* pangenome which contains the union of all genes that encode metabolic reactions from 16 genomes of *E. coli*, 2) using flux balance analysis to systematically test growth predictions in three simulated host environments of all single gene mutants, and 3) comparing the essential/important gene predictions (i.e. those that promote growth and would likely have been retained over time) with sequenced enterobacterial genomes to determine if these genes were retained or lost in modern day strains.

In this work, we have developed new methods using constraint-based optimization and metabolic model construction to identify genes important for growth/survival in environments simulating three locations within the human body and have compared the predictions with actual evolutionary outcomes of sequenced genomes of enterobacterial pathogens, such as extraintestinal *E. coli*, that cause human disease in locations other than the intestinal tract. Extraintestinal *E. coli* infections may result in serious illness and even death, and globally 130–175 million cases of urinary tract infections are caused by Extraintestinal *E. coli* [10]. The urinary tract is also the most common route for *E. coli* causing bloodstream infections, which cause more than 40,000 deaths from septicemia each year worldwide [10]. Therefore, an understanding of the genes that are essential for the growth of these pathogens to survive in certain human body niches is of great interest to aid efforts on developing new control strategies and therapeutics.

Computational modeling allows us to conduct experiments of disease-causing bacteria where actual testing in humans is not an option. These are the three main objectives that were investigated: i) Can different locations in the human body be modeled using constraint-based linear programming? ii) Are there different predictions of essential/important genes for growth in simulated conditions representing three human body locations? iii) Do these gene predictions correlate with the genome content of modern-day enterobacterial pathogens that actually

cause disease in each of the three locations? Overall, this study illustrated that mathematical constraints can be used with metabolic models to simulate the nutrient conditions the pathogen encounters during the infection process, and the genes predicted using FBA with the metabolic model simulating conditions during infection correlate with transcriptional gene-expression data obtained for conditions representing host-pathogen interactions. The central hypothesis is that the essential and important genes for bacterial growth in certain environments should be mostly remained over time in the genome of strains that cause disease in the corresponding human body locations, whereas the loss of those essential and important genes should not cause dire consequence for strains that invade different human locations.

Results and Discussion

Computational simulation of different niches in the human body

For the three simulated conditions, analytical data were used to add constraints that dictate metabolite availabilities respectively under three simulated conditions, the human macrophage cell [9], the bloodstream [11], and the urinary tract [12]. During the macrophage invasion, the pathogens can be engulfed and chained inside the pathogen-containing vacuoles that may restrict nutrients for cell growth. There is very little information on the nutrient compositions of those vacuoles under different macrophage activation states. Considering the pathogens may achieve nutrients from cytoplasm by modifying the membrane of vacuoles, existing literature values on the nutrient composition of the macrophage cytoplasm can be used to mimic the environment inside a macrophage for pathogen growth.

For the three simulated niches examined in human body, there were 15 available metabolites used as constraints shared in common for all three host niches, whereas 51 metabolites varied depending on the environment, indicating that differences in human body locations lead to different metabolite compositions available to the microrganisms (Table 1).

Predictions of essential/important genes for cell growth in three simulated human body locations

When FBA analysis for single reaction deletions and their corresponding genes was conducted in the three simulated environments, the results varied in the total number of predicted essential and important reactions and associated genes for each condition (Table 2). Following each gene deletion, if the rate of biomass production was calculated as a value of zero (no growth prediction) or a reduction of >1% of the wild type biomass production, the genes were considered to be essential or important, respectively. There were 38 reactions predicted to be commonly essential for all three simulated human body locations, as the absence of them led to no cellular growth (Fig 1). Besides, 38 reactions were predicted as essential that were not shared in common for those conditions (Fig 1). There was only one reaction predicted to be important that resulted in a decrease of predicted biomass for all three simulated host locations, whereas 121 reactions were predicted as important that led to a predicted biomass reduction in one or two simulated conditions (S1 Data). For all of these essential and important reactions the genes correspond to, the reactions were identified to report the number of essential or important genes' lost (S2 Data).

Comparison of essential/important gene predictions based on the genomes of real disease-causing enterobacterial pathogens in each of the three host niches

Once the essential and important genes were identified, they were compared with the sequenced genomes of enterobacterial pathogens that invade the macrophage cell, infect the

Table 1. Nutrients used to simulate three host environmental conditions.

Metabolites	Macrophage	Blood	Urine
2-Oxoglutarate	-	+	-
Acetoacetate	-	+	-
Adenine	-	-	+
Adenosine	-	+	-
Allantoin	+	+	+
Arabinose	+	-	-
Butyrate	-	+	+
Carnitine	+	-	-
Citrate	-	+	+
Cytosine	+	-	-
Deoxycytidine	+	-	-
Ethanolamine	+	-	+
Formate	-	-	+
Fructose	+	-	-
Fucose	+	-	-
Fumarate	-	+	-
Galactarate	+	-	-
Galactonate	+	-	-
Glucarate	+	-	-
Gluconate	+	-	-
Glucosamine	-	+	-
Glucose	+	+	+
Glucuronate	+	+	+
Guanine	-	-	+
Hypoxanthine	+	-	-
Inosine	+	-	-
D-lactate	-	+	+
L-lactate	-	+	+
L-Malate	-	+	-
D-Malate	-	+	-
Maltose	+	-	-
Mannitol	+	-	-
Mannose	+	-	-
Melibiose	+	-	-
Myo-Inositol	-	+	+
N-Acetyl-D-glucosamine	+	-	-
N-Acetylneuraminate	+	-	-
Nicotinate	-	+	-
Pantothenate	+	-	-
Propane-1,2-diol	+	-	-
Putrescine	+	-	-
Pyruvate	-	+	+
Rhamnose	+	-	-
Ribose	+	-	-
Sorbitol	+	-	-
Spermidine	+	-	-
Succinate	-	+	-

(*Continued*)

Table 1. (*Continued*)

Metabolites	Macrophage	Blood	Urine
Taurine	-	-	+
Thiamin	+	+	-
Uracil	+	-	-
Uridine	+	-	-

Present / Not Present = + / -

doi:10.1371/journal.pone.0149423.t001

bloodstream, or cause disease in the urinary tract. Three genomes (*E. coli* UTI89, *E. coli* 53638, and *Salmonella* LT2) were used for essential and important gene comparison, and the genome of *E. coli* O157:H7 was used as a control because of the pathogen's capability to cause disease in the human intestine. *E. coli* UTI89 is able to infect the urinary tract or the bloodstream in human body, causing disease outside the intestinal track. Both *E. coli* 53638 and *Salmonella* LT2 can cause disease by invasion of a host cell (Table 3).

The central hypothesis is that the pathogens that actually cause disease in a given host location should have lost the fewest number of essential and important genes predicted for that conditions simulated *in silico* (macrophage, bloodstream, or urinary tract). In contrast, the pathogenic *E. coli* O157:H7 that causes disease in the intestinal tract would most likely have lost the most number of essential and important genes predicted for each of the three host niches. The host niche condition was not simulated for the control in this project. As shown in Table 4, when compared to the genomes of these organisms, the number of lost essential and important genes in each strain varied. When the numbers of both lost predicted essential and important genes out of the total number are summarized (Table 5), it is clear that some of the predictions match the real evolutionary outcomes of the genome content of these organisms, whereas the simulation of the urinary tract did not match the evolutionary outcomes of these strains, and this discrepancy is addressed in the conclusions section.

Conclusions

This study investigated *in silico* metabolic modeling and prediction of genes required for growth and survival in three human body locations. Based on the numerous differences of metabolites present in three different human body niches, this study illustrates that multiple environmental niches in a human can be simulated to study microbial metabolism by using constraint-based linear programming and computational model. Simulation of these three conditions led to different predictions of essential and important genes/reactions, which match the real evolutionary outcomes when compared to the control genome of the intestinal pathogen enterohemorrhagic *E. coli* O157:H7 strain EDL933, a strain that causes disease in the intestine and was predicted to have lost the most of the essential or important genes in the three other host niches. In the case of intracellular invasion, although the strain isolated from a

Table 2. Total number of reactions and corresponding genes predicted as essential and important for growth in three simulated human body locations.

Host niche	Essential reactions	Important reactions	Essential genes	Important genes
Macrophage	195	146	290	146
Bloodstream	193	65	288	182
Urinary tract	203	52	304	151

doi:10.1371/journal.pone.0149423.t002

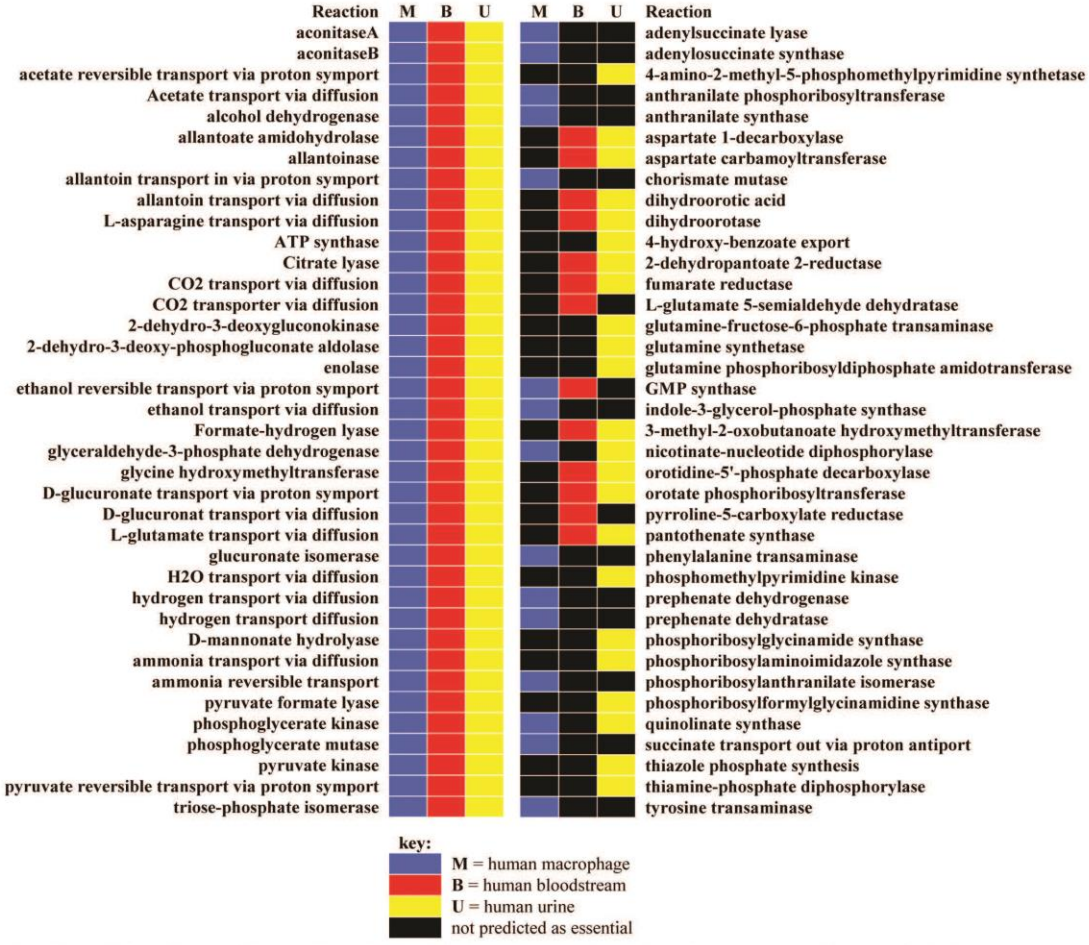

Fig 1. Essential reactions predicted for three simulated host environmental conditions. There are 38 reactions predicted to be commonly essential for all three simulated human body locations, whereas 38 essential reactions predicted that are differed for simulations of the human bloodstream, urinary tract, and macrophage.

doi:10.1371/journal.pone.0149423.g001

Table 3. *E. coli* and *Salmonella* genomes used in this study.

Host niche	Enterobacterial human pathogenic strains	Genome of strain that causes disease
Bloodstream	Extraintestinal pathogenic *E. coli*	*E. coli* UTI89
Macrophage	*Salmonella* spp., Enteroinvasive *E. coli*	*E. coli* 53628, *Salmonella* LT2
Urinary tract	Urinary tract pathogenic *E. coli*	*E. coli* UTI89
Intestinal tract (control)	Enterohemorrhagic *E. coli*	*E. coli* EDL933

doi:10.1371/journal.pone.0149423.t003

Table 4. Total number of predicted essential and important genes lost out of total predicted for each strain.

Host Niche	Genes Lost/Total Predicted	*E. coli* 53638	*E. coli* UTI89	*Salmonella* LT2	*E. coli* O157:H7
Macrophage	essential genes	2/290	4/290	3/290	13/290
Macrophage	important genes	20/366	17/366	22/366	58/366
Bloodstream	essential genes	2/288	1/288	3/288	12/288
Bloodstream	important genes	14/182	12/182	18/182	26/182
Urinary tract	essential genes	2/304	6/304	4/304	12/304
Urinary tract	important genes	9/151	11/151	15/151	19/151

doi:10.1371/journal.pone.0149423.t004

urinary tract infection has the fewest essential/important genes lost, the two genomes of strains that actually cause disease through this route had very similar low numbers of lost essential/important genes. In the case of the simulations for the human bloodstream and urinary tract, *E. coli* UTI89 is the strain that actually causes disease in these locations, and had the least amount of necessary/important genes lost, which agreed with the evolutionary outcome. This project demonstrated that human body environments such as the bloodstream can successfully lead to accurate predictions of essential/important genes using optimization and constraint-based metabolic techniques. The discrepancies from the predictions for the urinary tract may indicate that more information is required for additional constraints to more accurately simulate this environment, or that the *E. coli* strains that have been characterized as causing disease in only one niche in the human body may also be capable of causing disease in numerous locations in the human body. Overall, this project was a success and lays a foundation towards future work to model metabolism of pathogenic microbes in different locations inside a human host. Since the actual infection study of these organisms in human is not a possibility, computer modeling of related disease processes becomes an emerging approach and field that is likely to grow immensely. By addressing these research ideas revealed by this project using optimization and constraint-based linear programming, the field of microbial system biology can be furthered to efficiently examine genome evolution.

Materials and Methods

Pangenome Metabolic Network Reconstruction

The metabolic model representing the *E. coli* pangenome (iEco1712_pan) used in this work was previously reconstructed based on the gene to protein to reaction (GPR) information of 16 *E. coli* genomes obtained from the ASAP database [1]. Draft and complete genomes have been continually updated using new publicly accessible genomes in the ASAP database since its inception [13]. There currently are 39 genomes among more than 150 enterobacteria genomes in the ASAP database that belong to *E. coli*, of which 16 are completely finished and were used

Table 5. Total number of predicted essential and important genes lost out of total predicted for each strain.

Host niche	*E. coli* 53638	*E. coli* UTI89	*Salmonella* LT2	*E. coli* EDL933 (control)
Macrophage	22/656[c]	21/656[c]	25/656[c]	71/656[a]
Bloodstream	16/470[a]	13/470[a]	21/470[a]	38/470[a]
Urinary tract	11/455[b]	17/455[b]	19/455[b]	31/455[a]

[a]Evolutionary outcome agrees with *in silico* predictions for genome content

[b]Evolutionary outcome disagrees with *in silico* predictions for genome content

[c]Evolutionary outcome is within standard deviation with *in silico* predictions for genome content

doi:10.1371/journal.pone.0149423.t005

in the construction of the metabolic model of *E. coli* pangenome (iEco1712_pan) [1]. The reconstructed network contains metabolic enzymes present in a union of 76,080 Open Reading Frames (ORFs) that map 17,647 Clusters of Orthologous Groups (COGs), with each ORF being assigned to an COG in the ASAP database, and all of the information for model composition, GPR associations for the *E. coli* pangenome (iEco1712_pan) reconstruction used in this work are available as supplemental information along with the sbml file for the iEco1712_pan GEM [1].

Flux Balance Analysis

Flux balance analysis (FBA) has been commonly applied for mathematical analysis of GEMs, which can predict reactions-related fluxes in a metabolic network [14]. By constraining fluxes with steady-state mass balances, reaction directionality, and metabolite availability, a range of possible flux values can be generated in FBA. An objective function then can be used to identify flux distributions that maximize (or minimize) the objective function with those constraints. Biomass production, a commonly used objective function for FBA performance and for a proxy of growth, was adapted in this study [15]. FBA was conducted using the software package GAMS in this study, in which the *E. coli* pangenome metabolic network is described as a stoichiometric matrix ($\mathbf{S_{i,j}}$) with rows (i∈I) representing the metabolites and columns (j∈J) indicating reactions that correspond to genes (g∈G). In a steady-state, the mass balance equation can be described as below, with v being the flux vector. Additional constraints are showed as lower and upper limits for the values of fluxes through reactions in a network.

$$\begin{aligned} &Max\ Vbiomass \\ &\text{s.t. } S_{ij} \bullet \mathbf{v} = 0 \\ &v_{j,lb} < v_j < v_{j,up} \end{aligned} \tag{1}$$

The matrix built for *E. coli* pangenome GEM contains 1,726 metabolites (I) and 2,324 reactions (J) that associate with 1,712 genes (G). Three different niches located in the human body (macrophage, blood, and urinary tract) were simulated to set constraints for FBA in this study, with possible metabolite compositions being identified through literature review that determined analytical compositions of nutrients present in each bodily location. The simulated condition for macrophage contains 32 metabolites, the bloodstream environment contains 19 metabolites, while there are 14 metabolites that belong to the urinary tract niche (Table 1).

Gene Essentiality

Unlike virulence factor genes [16], essential genes are those required to maintain critical cellular functions under specific environments, while important genes are not irreplaceable but still necessary for robust bacterial growth under those conditions. To determine the essentiality of genes expressed under different environmental pressures (macrophage cell, bloodstream, and urine tract), genes were removed one-by-one in networks and the resulting changes in biomass production rate can be estimated to reveal the impact of gene loss (a proxy for fitness). Following each gene deletion, if the calculated value of biomass production rate was zero, meaning no predicted intracellular growth, the gene would be considered essential. Important genes were predicated based on >1% reduction of the wild type biomass production rate. A graphic description on identifying essential genes and corresponding metabolic reactions using GEMs constructing and computational predictions is showed in Fig 2.

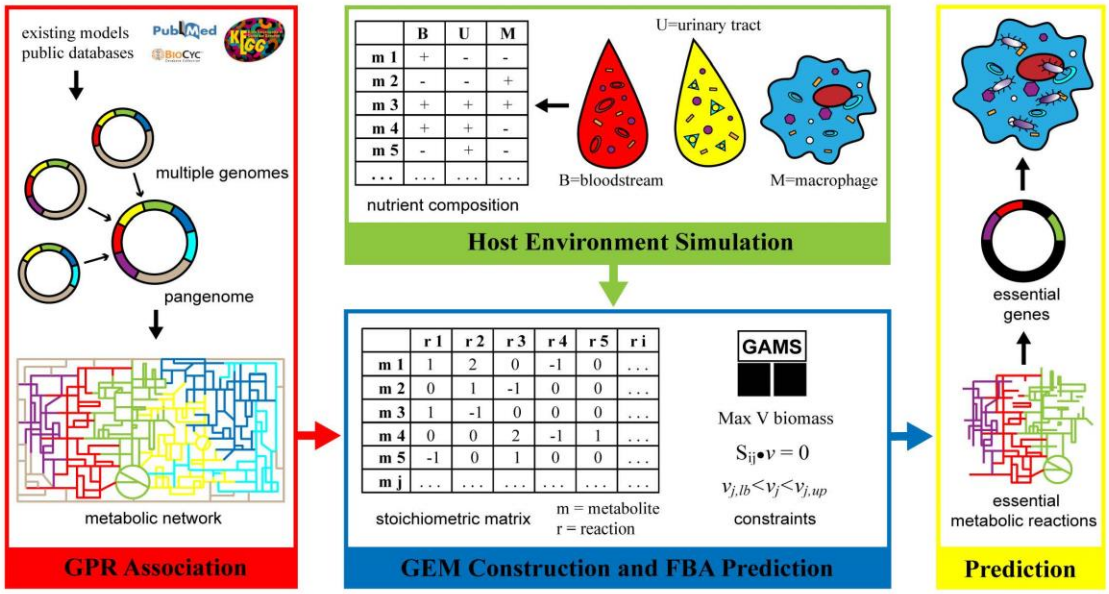

Fig 2. Essential gene identification using GEMs predictions under simulated environment. The GEM constructed upon pangenome incorporated from *E. coli* genomes can be used to generate predations with simulated nutrient conditions to identify essential genes along with corresponding essential metabolic reactions under multiple human body niches.

doi:10.1371/journal.pone.0149423.g002

Supporting Information

S1 Data. Reactions predicted as important for all three simulated environmental conditions and that differed during simulation of the human bloodstream, urinary tract, and macrophage.
(XLSX)

S2 Data. Reactions corresponding to essential gene predictions for the *E. coli* pangenome GEM. This file contains three tables, the first contains all predicted essential reactions during simulation of human macrophage, the second contains all human bloodstream predicted essential reactions, and the third contains predicted essential reactions during simulation of the human urinary tract.
(XLSX)

Acknowledgments

We would also like to thank Dr(s). William R. Harcombe, Guy Plunkett III, Bob Mau, and Eric Cabot for insightful discussions regarding gene essentiality within the genomes of members of the family *Enterobacteriaceae*. This work was partially funded by the Department of Food Science and Nutrition and the College of Food, Agricultural and Natural Resource Sciences at the University of Minnesota-Twin Cities (DJB), and partially by the Global Food Ventures Graduate Student Fellowship from the University of Minnesota-Twin Cities (ZPM), and the Schlumberger Faculty for the Future Graduate Student Fellowship (MAO).

Author Contributions

Conceived and designed the experiments: DB. Performed the experiments: TD KC MO HR ZM XD. Analyzed the data: TD KC MO HR ZM XD DB. Wrote the paper: TD KC MO HR ZM XD DB.

References

1. Baumler D, Peplinski R, Reed J, Glasner J, Perna N (2011) The evolution of metabolic networks of E. coli. BMC Systems Biology 5: 1–21.
2. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3: 121. PMID: 17593909
3. Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol 26: 659–667. doi: 10.1038/nbt1401 PMID: 18536691
4. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. (2013) Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences 110: 20338–20343.
5. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. Mol Syst Biol 7: 535. doi: 10.1038/msb.2011.65 PMID: 21988831
6. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology 4: R54–R54. PMID: 12952533
7. Baumler DJ, Ma B, Reed JL, Perna NT (2013) Inferring ancient metabolism using ancestral core metabolic models of enterobacteria. BMC Syst Biol 7: 46. doi: 10.1186/1752-0509-7-46 PMID: 23758866
8. Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. Nature 440: 667–670. PMID: 16572170
9. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. BMC Syst Biol 3: 38. doi: 10.1186/1752-0509-3-38 PMID: 19356237
10. Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect 5: 449–456. PMID: 12738001
11. Keitel HG, Berman H, Jones H, Maclachlan E (1955) The Chemical Composition of Normal Human Red Blood Cells, including Variability among Centrifuged Cells. Blood 10: 370–376. PMID: 14363319
12. Putnam DF (1971) Composition and concentrative properties of human urine. McDonnell Douglas Astronautics Company. #CR-1802. 1–112 p.
13. Glasner JD, Rusch M, Liss P, Plunkett G, Cabot EL, Darling A, et al. (2006) ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Research 34: D41–D45. PMID: 16381899
14. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nature biotechnology 28: 245–248. doi: 10.1038/nbt.1614 PMID: 20212490
15. Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol 13: 344–349. doi: 10.1016/j.mib.2010.03.003 PMID: 20430689
16. Reiland HA, Omolo MA, Johnson TJ, Baumler DJ. (2014) A survey of Escherichia coli O157:H7 virulence factors: The first 25 years and 13 genomes. Advances in Microbiology. 4:7.

Chapter III Microbial Risk Analysis of Produce Grown On a Sustainable Chicken Production Farming System

III. 1 **Introduction**

Sustainable agriculture is defined as "an integrated system of plant and animal production practices having a site-specific application" by USDA National Institute of Food and Agriculture [56]. To perform sustainable farming practices, organic fertilizer, which is a soil amendment derived from natural sources, such as animal by-products, has been widely used. A well-designed sustainable faming system should lead to lower energy costs, environmental preservation, and many other benefits. However, the system designed for one specific farm may not work for another, and; therefore, the unique features of the farm must be considered when designing a system to fit that farm [57]. In this study, the farming system was designed for property located in Northfield, MN, possessing free range chickens, perennial plantings of hazelnuts and elderberries, and annual vegetable crops.

The organic fertilizer used for sustainable farming was poultry manure, a good source of nutrients that foster the growth of crops, but also a natural reservoir of human pathogens such as *Salmonella* [58-60]. For the purpose of energy-saving and financial benefits in a sustainable system, the manure had been collected from a local meat broiler and stockpiled outside throughout the winter period. With temperatures that can fall as low as −60 °F (−51 °C) and heavy snowfalls (2.3 to 170 inches on average) [61], this stockpiling process was integrated within the farming system to inhibit the growth and multiplication of potential pathogens in the manure fertilizer under the extremely cold winter climate in Minnesota.

To evaluate the biological safety of this sustainable farming system, microbiological hazards must be identified, because pathogens present in the air, water, soil, animals, and animal feces can cause microbial contamination that may result in foodborne illness outbreaks. Coliforms, including *Escherichia coli* (*E. coli*) that was found in the gut flora of warm-blooded animals, are commonly used as indicator microorganisms to predict the level of fecal contamination in food and water [62]. In this study, the fecal coliform *E. coli* was tested considering its pathogenicity related to human illness. The presence of *Salmonella* was also tested, because it is widely found in poultry products and can inhabit a wide range of niches from warm-blooded animals to plants [63]. *Listeria monocytogenes* was also tested, due to its ability to survive and multiply under refrigeration temperature [64]. Chickens can be asymptomatic disease-carriers for these bacteria, and the pathogens can contaminate fresh produce through fecal contamination. Reused chicken manure can possess up to 9.7×10^4 CFU/g *E. coli*[58]. The resulting foodborne illnesses pose a threat to high-risk population including young, old, pregnant, and immuno-compromised individuals [65].

III. 2 **Materials and Methods**

In this study, microbial risk was analyzed by detecting the presence or absence of *Salmonella*, *Listeria*, and coliforms including *E. coli* in samples received from a sustainable farming test field that uses chicken manure as soil fertilizer. According to the farm worker, the poultry litter used in this test field was harvested from a meat broiler unit at Mirasol Farm in October 2014, applied in a 3 inch thick layer, and left between

elderberry rows throughout winter. In spring 2015, the manure was turned lightly and applied to crops as fertilizer. Starting in May 2015, one bag of fertilized soil and one bag of unfertilized soil were collected by the farm worker and delivered to our Biosafety Level 2 Laboratory for microbial detection until September. In addition to the comparison between fertilized and unfertilized soils, tests were also done for the chicken manure sample collected in May, the spinach sample collected in June, and the cantaloupe samples collected in August to better interpret the analysis result.

Salmonella detection was done qualitatively over the 5 months based on the current FDA Bacteriological Analytical Manual (BAM) [66]. 3M™ Coliform Count Plates were used for quantitative coliform detection in May, and then 3M™ *E. coli*/coliform Count Plates were used instead until September, so the presence or absence of *E. coli* can be detected as well. One detail not included is this type of petrifilm does not specifically indicate whether any O157 strain is present. 3M™ Petrifilm™ Environmental *Listeria* Plates were used for qualitative *Listeria* spp. detection from June to September, after a decision was made to test *E. coli* and *Listeria* other than *Salmonella* and coliforms. Therefore, the manure sample collected in May was not detected for *Listeria* spp. and *E. coli*. 0.1% peptone water was used in serial dilution that was done in duplicate for petrifilm inoculation without enrichment, and the average amount of detected bacteria was calculated for each sample.

III. 3 **Results and Discussion**

Coliforms commonly inhabit the intestinal tract of chickens, and; therefore, they can be found in chicken litter and soil that was contaminated by the manure. By comparing the amounts of coliforms in normal soil, manure-fertilized soil, chicken manure used as soil amendment, and fresh produce cultivated in the amended soil, the potential of fecal contamination caused by applying poultry manure as organic fertilizer can be estimated. Table 4 shows detailed counts of coliforms and *E. coli* detected in normal soil, amended soil, and fresh produce samples over five months. On average, the coliform population detected in the samples of manure-fertilized soil ($1.53\times10^4 \pm 4.26\times10^3$ CFU/g) is slightly higher than the amount of coliforms found in the unfertilized soil samples ($1.37\times10^4 \pm 4.77\times10^3$ CFU/g), while the pure poultry manure that was used as soil amendment contains 5.50×10^5 CFU/g coliforms.

Table 4: The amount of coliforms/*E. coli* (CFU/g) quantitatively detected in fertilizer, manure amended soil, unfertilized soil and fresh produce samples over five months

	May	**June**	**July**	**Aug.**	**Sept.**
Fertilized soil[1]	9.75×10^3/ N.A.[2]	1.93×10^4/ BDL[3]	3.75×10^3/ BDL	5.10×10^4/ BDL	8.38×10^3/ 7.50×10^1
Unfertilized soil[4]	4.00×10^2 / N.A.	8.50×10^3/ BDL	4.98×10^4/ BDL	1.05×10^4/ 5.00×10^1	8.13×10^2 / BDL
Spinach	– [5]	2.35×10^4/ BDL	–	–	–
Cantaloupe 1	–	–	–	5.00/ BDL	–
Cantaloupe 2	–	–	–	3.10×10^2/ BDL	–
Fertilizer[6]	5.50×10^5 / N.A.				

[1] Soil amended with poultry litter, used for planting fresh produce
[2] Not Applicable – no test was performed
[3] Below Detection Limit – the presence of colonies was not observable on test
[4] Soil without chicken manure amendment, collected from the same farm
[5] No sample of such type was collected for the test
[6] Organic fertilizer (poultry manure) used for crop cultivation, tested only once

The number of coliform counts detected in fertilized soil fluctuated around the average value over the five months as shown in Figure 5, while for unfertilized soil the number increased from May to July and then dropped till September. This trend may relate to the climate changes over the five months, in which the temperature and moisture level picked around July [67]. The fertilized soil probably kept a consistent moisture level over that period due to irrigation compared to the unfertilized soil that was not used for farming, so its microorganism content was relatively stable. On a log scale, the figure better indicated that the coliform populations detected in the samples of manure-fertilized soil and unfertilized soil samples were not statistically different compared to the amount of coliforms found in the poultry manure sample.

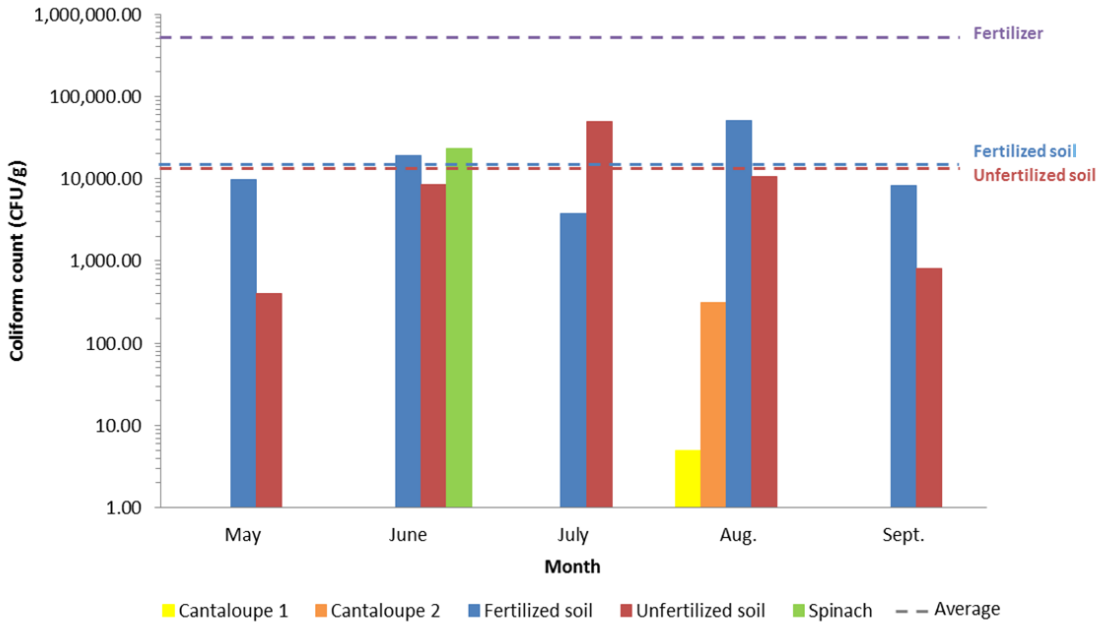

Figure 5: The amount of coliforms (CFU/g) quantitatively detected in fertilizer, manure amended soil, unfertilized soil and fresh produce samples over five months.

Both cantaloupe and spinach grow outside on the ground, exposing themselves to soil, rain, and wild animals like birds; therefore, they are at great risk for biohazards

related to contaminated environment. In 2006, a multistate outbreak of *E. coli* O157:H7 infections occurred and the source was identified to be fresh spinach [68]. The number of coliforms detected on the spinach sample collected in June was higher than the amount of coliforms found in the amended soil in the same month, but no *E. coli* was found in both samples. The two cantaloupe samples tested in August also do not contain detectable *E. coli*, and the amounts of coliforms they possess were much lower than the value estimated for the fertilized soil sample.

Similarly, the experimental results indicated that *Salmonella* should not be a concern for planting crops using chicken manure as fertilizer in this farming system, as there was no *Salmonella* qualitatively detected on either manure-treated soil or fresh produce samples. In contrast to the *Salmonella* results, the presence of *Listeria* was verified in the unfertilized soil sampled in the last three months of the study, and it was also found in one of the cantaloupe samples in August (Table 5).

Table 5: The presence of Salmonella/Listeria qualitatively detected in manure amended soil, unfertilized soil, fertilizer and food samples over five months

	May	**June**	**July**	**Aug.**	**Sept.**
Fertilized soil[1]	N[2]/N.A.[3]	N/N	N/N	N/N	N/N
Unfertilized soil[4]	N/N.A.	N/N	N/Y[5]	N/Y	N/Y
Spinach	–[6]	N/N	–	–	–
Cantaloupe 1	–	–	–	N/N	–
Cantaloupe 2	–	–	–	N/Y	–
Fertilizer[7]	Y/Y				

[1] Soil amended with poultry litter, used for planting fresh produce
[2] Not qualitatively detected
[3] Not Applicable – no test was performed
[4] Soil without chicken manure amendment, collected from the same farm
[5] Qualitatively detected
[6] No sample of such type was collected for the test
[7] Organic fertilizer (poultry manure) used for crop cultivation, tested only once

The Petrifilm used for *Listeria* detection in this study detects environmental *Listeria* spp. including *L. monocytogenes*, *L. innocua*, and *L. welshimeri*. Although the *Listeria* spp. found in the samples of this study cannot be further identified, the presence of *Listeria* spp. provides evidence that the environment is suitable for the occurrence of *L. monocytogenes*. This pathogen is tolerant to refrigeration temperature [69], and it is possible for them to survive under the winter climate in Minnesota. Cantaloupes contaminated with *Listeria monocytogenes* has caused an multistate outbreak of listeriosis in 2011 [70].

Although previous studies have stated that *L. monocytogenes* should not be a concern for chicken manure-based organic fertilizers, considering it is usually absent from chicken waste [58], due to the limitations in sampling we cannot identify the source of *Listeria* contamination. As the unfertilized soil samples showed positive in *Listeria* presence, it is possible that the cantaloupe was contaminated by factors other than the use of manure-amended farming soil. Practices such as soil amendment application, irrigation before harvest, gathering, handling, and processing after harvest are factors that may influence the microbial safety of the fresh produce. Climate change, wildlife interference, and geographical location can also affect the safety of food.

The stockpiling process utilized in this sustainable chicken production farming system is intended to suppress the growth of pathogens with cold temperature throughout the winter in Minnesota. Nevertheless, small amounts of bacteria may regenerate from a small population after surviving temperature-dependent processes when they encounter suitable conditions [71]. Therefore, the time interval between raw manure application and

harvest has an impact on the risk of crop contamination. For crops in contact with the soil, the FDA Food Safety Modernization Act proposed an application interval of nine months but later compromised with a 120-day interval for the farmers complying with the USDA's National Organic Program standards [72]. An interval of at least one year for the use of raw manure is suggested by The Leafy Greens Marketing Agreement, considering the long survival period of bacteria in raw manure that poses a risk for lettuce and leafy greens [73].

As Chen *et al*. suggested, subsequent treatments should be added to more efficiently inactivate pathogenic microorganisms in chicken litter [58]. Additional control parameters, such as heat or chemical treatment, are recommended to minimize the risk of contamination and improve safety [60, 71]. Field experiments by Nicholson *et al*. showed that if the temperature in solid manure heaps surpasses 55 °C, the level of *E. coli*, *Salmonella*, and *Listeria* will be undetectable within one week [74]. Destruction of the three microorganisms can also be greatly increased by decreasing the moisture content and exposing the manure to ammonia gas [75].

Overall, this work found the food samples produced in the sustainable farming system are safe from *E. coli* and *Salmonella*. Although there currently is no standard for the *Listeria* on the surface of cantaloupes, it is suggested that additional biohazard control approaches should be implemented into the practice of sustainable farming to improve the safety of produce items.

Chapter IV Analysis on the Metabolic Capabilities of five *Salmonella* Strains through Genome-scale Metabolic Models

IV. 1 **Introduction**

Food related *Salmonella* strains are mainly non-typhoidal, and this type of *Salmonella* is ranked the number one human food poisoning risk among all the bacterial pathogens tracked by the CDC. Unlike typhoidal *Salmonella* that is problematic mostly in developing areas with sanitation issues, non-typhoidal *Salmonella* is more of a global concern as it leads to about 93.8 million illnesses and 155,000 deaths annually throughout the world. *Salmonella* infections are usually self-limiting in healthy adults, while the high-risk population including children and the elderly may suffer from prolonged complications such as bacteremia [5]. The clinical treatment for such complications is limited to antibiotic therapy, and unfortunately some *Salmonella* strains have already developed resistant capabilities against antimicrobials due to the overuse of antibiotics. For example, the emerging *Salmonella* serovar 4,[5],12:i:- used in this study is resistant to multiple antibiotics including ampicillin, chloramphenicol and streptomycin [33]. The emergence of these drug-resistant strains has become a huge public health concern, and therefore we need to develop new strategies and targets of control against *Salmonella* infections.

For the last few decades, the focus on virulence factor genes has been the attention of a great deal of research, but recently there is immense interest in the interplay between bacterial metabolism and virulence pathways of intestinal pathogens such as *Salmonella* Typhimurium and *Vibrio cholera* [76]. During bacterial evolution, many

microorganisms either obtained new metabolic genes or lost genes that encode unnecessary metabolic properties so they can thrive in a new environment or better compete with other bacteria in the same environment. In the case of *Salmonella* Typhimurium, the organism has gained unique genomic islands that confer new metabolic capabilities for utilization of ethanolamine and 1,2-propanediol as carbon sources found in the intestinal tract through a unique tetrathionate reduction pathway. These acquired metabolic capabilities confer *S.* Typhimurium a competitive advantage in the human intestine, since no other gut microbes can generate energy from these substrates and tetrathionate inhibits the growth of coliform, thus allowing *S.* Typhimurium to grow faster and outcompete the intestinal normal flora [76].

A bacterium's pathogenicity largely depends on their metabolic capabilities during host infection or environmental adaptation, and therefore by gaining a better understanding on bacterial metabolism it is possible to develop new target for potential pathogen control [77]. The genome of a bacterium contains essential genes that encode a number of basic metabolic reactions functioning to assist cell survival [78], and by assembling the corresponding gene to protein to reaction associations the metabolic network can be reconstructed *in silico* as a genome-scale metabolic model (GEM). This systems biology approach has been proved to be ideal for studies on complex biological networks, as it is able to integrate and process high throughput genomic and proteomic data to conduct qualitative as well as quantitative analysis on bacterial metabolic properties. Aside from the network analysis, other applications of GEMs include metabolic engineering, biological discovery of potential reactions, and comparison of

bacterial evolution patterns based on a contextual network background present by the GEM [49]. Up to date, more than 50 of GEMs have been published for various microorganisms including *Yersinia pestis*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Shewanella oneidensis*, and *E. coli* [79-83]. Among the many disease-causing *Salmonella* serovars and strains, only the serovar Typhimurium strain LT2 has been studied through the use of GEM [43, 47, 48].

The goal of this study was to construct GEMs for five different strains under four representative *Salmonella* serovars through a semi-automatic modeling approach. Serovar Typhimurium (str. UK1 and str. LT2) and Enteritidis (str. P125109) are the most prevalent causes of human food poisoning and animal infections. *S.* Typhimurium can be found in a wide range of hosts from vegetables to warm-blooded animals, while most outbreaks caused by *S.* Enteritidis relate to poultry and egg products. The serovar 4,[5],12:i:- (str. CVM23701) has drawn more attention recently because of its multidrug resistance and infections linked to pork and the swine industry in Europe and more recently the United States. Serovar Abaetetuba (str. ATCC35640) was isolated from natural environment in 1983 and has not been recorded causing outbreaks in the U.S. The different preferences for host make these serovars ideal representatives to study differentiating metabolic capabilities among different microorganisms.

Each strain's GEM consists of a stoichiometric metabolic network preliminarily built by Knowledgebase (KBase), an open collaborative platform founded by the U.S. Department of Energy's division of Systems Biology for computational systems biology analysis of microorganisms. To make predictions through mathematical optimization, the

General Algebraic Modeling System (GAMS) software package was used to run flux balance analysis on the constructed models [84]. Additionally, experimental data on phenotypic nutrient utilization and quantitative bacterial growth were obtained for model validation, so the GEMs can be manually curated to improve the accuracy of their predictions. Using these GEMs, the metabolic differences between chosen strains can be analyzed to identify target genes/reactions in an integrative manner for the development of new pathogen control strategy and to decipher microbial host interactions related to niches in different types of animals.

IV. 2 **Methods and Materials**

Chosen Strains – Among the five strains chosen for GEM construction in this analysis, strain LT2, UK1, and ATCC35640 were available in the Baumler Lab culture collection for wet-lab experiments. These strains were restored from frozen stocks kept at -80°C and re-streaked onto Tryptic Soy Agar (TSA) plates once every two weeks. The M9 medium was selected as the chemically defined minimal medium for bacterial batch growth experiment, and the growth assay was performed on 96 well plates with the BioTek® Epoch 2 Microplate automated spectrophotometer to test the strains' viability on the M9 medium. The wells were inoculated and incubated at 37°C with continuous shaking for 48 hours while the OD_{600} readings were taken every 10 minutes.

GEM Construction – With the genome information collected from NCBI genbank, a preliminary GEM was generated for each of the chosen strains through the semi-automatic modeling platform KBase (Figure 6 A). The *in silico* metabolic network

reconstruction began with genome annotation using SEED functional roles based on the RAST algorithm [85]. The default annotation pipeline first annotated protein-encoding genes with k-mers (amino acid 8-mers) in CoreSEED, a database of ~1,000 microbial genomes. Then, PubSEED, another database that consists of ~12,000 microbial genomes, was used with k-mers to annotate the missed genes and corresponding proteins. Any remaining hypothetical proteins were annotated by searching against close relative genomes through BLAST and BLASTP. After that, the reconstructed networks were gapfilled with a simple linear programming formulation that identifies a minimal set of reactions that, when added to the model, will allow it to predict biomass generation on media where the microorganism is capable of growing on experimentally [86].

The reactions are written as equations in the computational models, and the metabolic network is expressed as a stoichiometric matrix (S) with rows representing metabolites and columns representing reactions (v). The numbers in the matrix are the stoichiometric coefficients for the metabolites participating in the corresponding reactions (Figure 6 B). With a mathematically defined model, flux balance analysis can be conducted in optimization software such as GAMS to estimate the predicted fluxes across the metabolic network through reactions. Assuming the network in question is at a steady state, meaning the amount of fluxes entering the system equals the amount leaves, a mass balance constraint can be set as $S \bullet v = 0$. Along with the constraints on substrate uptake that set the upper and lower bounds for fluxes ($a_i < v_i < b_i$), an allowable solution space can be defined. Then by optimizing an objective function, which is the biomass generation in this study, a single optimal solution can be identified such as qualitative and

quantitative growth predictions which can be compared to wet lab experimental data (Figure 6 C).

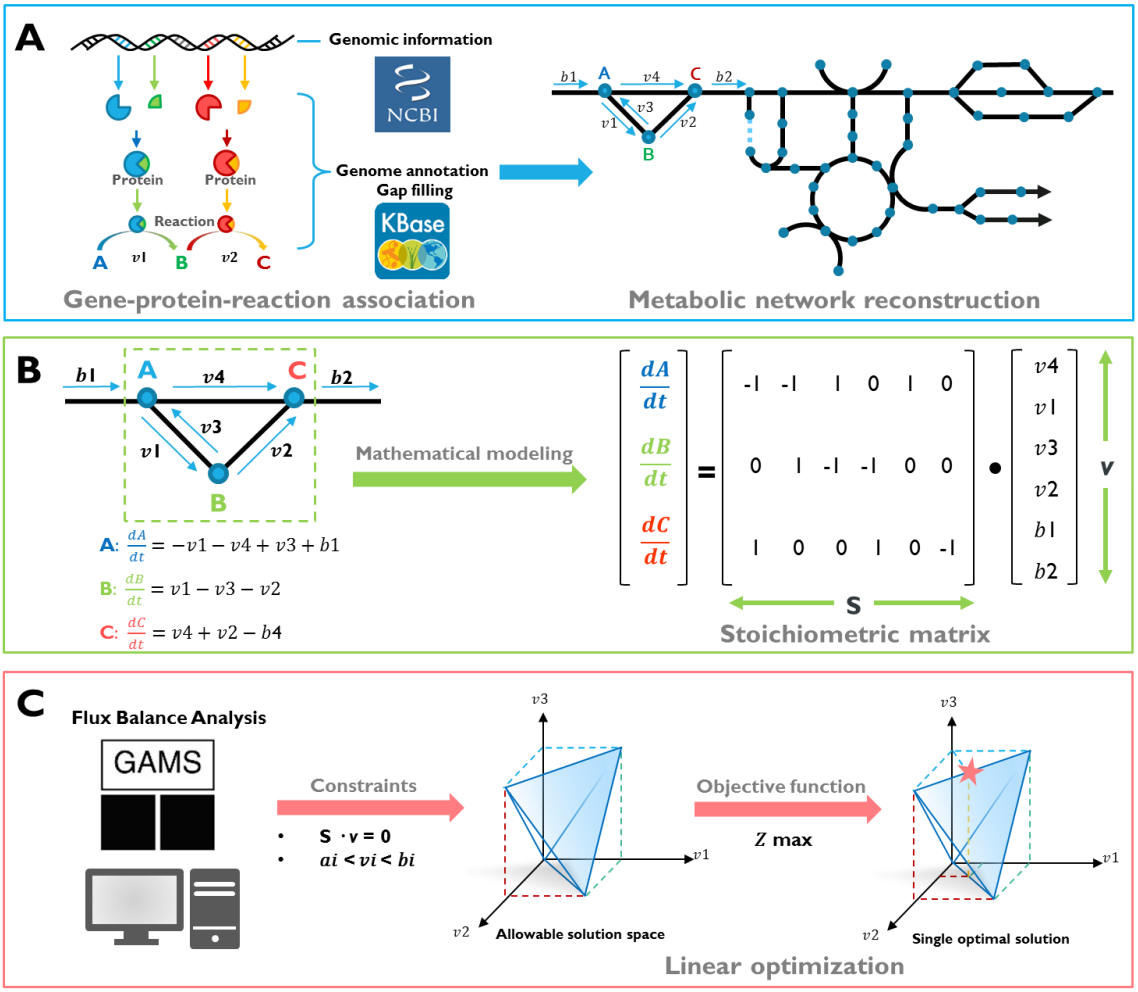

Figure 6: Reconstruction of *in silico* metabolic network based on gene-protein-reaction associations (A) and mathematical modeling (B) to perform flux balance analysis (C)

Qualitative Validation – Experiments were conducted using the Biolog™ Phenotype Microarray (PM) for bacterial nutrient utilization on carbon (PM1), nitrogen (PM3), phosphorous (PM4), and sulfur (PM4) substrates separately (Figure 7). The experiment started with culture enrichment on Biolog Universal Growth Agar with Blood (BUG+B)

under aerobic condition for PM1, PM3 and PM4 and anaerobic condition for PM1 at 37°C overnight. Then the bacteria lawn was harvested into conical tubes containing freshly made IF-0 solution (100 mM NaCl, 30 mM triethanolamine-HCl, 5.0 mM NH4Cl, 2.0 mM NaH2PO4, 0.25 mM Na2SO4, 0.05 mM MgCl2, 1.0 mM KCl and 1.0 µM ferric chloride) and its optical density was adjusted using IF-0 with the Spectronic GENESYSTM 20 Visible Spectrophotometer to receive a log phase cell suspension (OD_{600} = 0.171 ± 0.020). This cell suspension was transferred into the IF-0+ solution (IF-0 with 0.01% tetrazolium violet) before it was distributed into wells of Biolog™ PMs for incubation at 37°C.

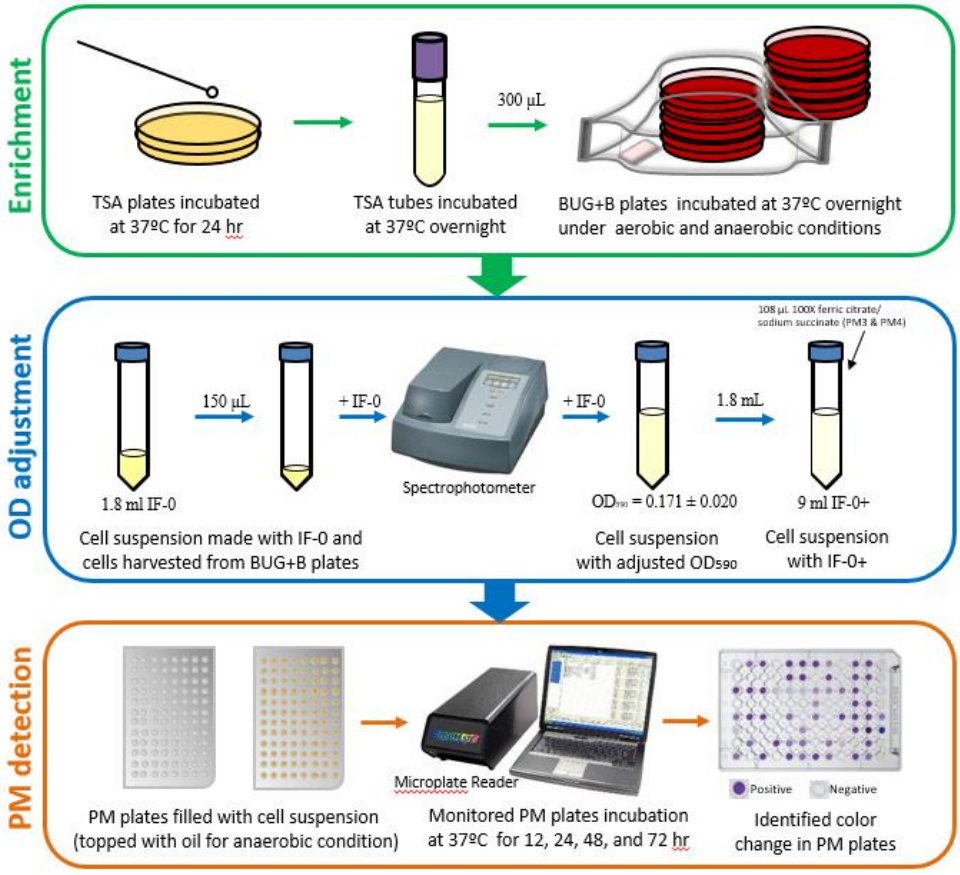

Figure 7: The flowchart of Biolog™ PM experiment for qualitative model validation

For PM3 and PM4 plates, 100X ferric citrate/sodium succinate solution was mixed into the suspension before distribution as additional nutrients that are essential for cell growth. For anaerobically incubated PM1 plates, each well was topped with low viscosity mineral oil sterilized through filtration (0.22 μm) after the cell suspension was distributed. A ChroMate® Microplate Reader was used to detect bacterial growth after 12, 24, 48 and 72 hours. Each well of the PM plate is designed to test a specific phenotype for a strain on nutrient utilization. If bacteria respire actively, the tetrazolium dye in the well will be reduced to form a strong color, showing a positive result in that well. The resulting bacterial phenotype information was used to compare with computational model predictions.

Quantitative Validation – The bacterial batch growth experiment was conducted by incubating M9 broth with bacteria culture and measuring OD_{600} as well as dry cell weight (DCW) as shown in Figure 8. To harvest sufficient amount of cells for DCW measurement, batch spargers were used as described in a previous study [87]. Each sparger bottle is an isolated incubation chamber that contains 600 mL of M9 minimal medium for bacterial growth either under aerobic or anaerobic conditions depending on the gas dispersed into the broth. To achieve a log phase OD_{600} in each sparge bottle, the amount of overnight enriched culture to be added was calculated so the resulting OD_{600} of cell suspension in each sparge bottle reached 0.040 ± 0.010 at the beginning of batch incubation. After the bottles were inoculated, the spargers were set up in a water bath at 37° C and cell suspensions were collected from each bottle for hourly OD_{600} measurement.

When the OD_{600} value reached 0.100, cell suspension was harvested from each sparger and filtered through pre-weighed dry Whatman™ glass microfiber filters. These filters were dried overnight in an 80°C oven one day before the experiment and they were dried again in the same condition for 24 hours after the experiment. The filtration step was repeated hourly three more times to make a total of four DCW measurements. After drying, the filters containing dried cells were weighed and the DCW determined by taking off the pre-weight value from the corresponding post-weight value. Biomass generation curves were made based on the experimental data to compare with model predictions.

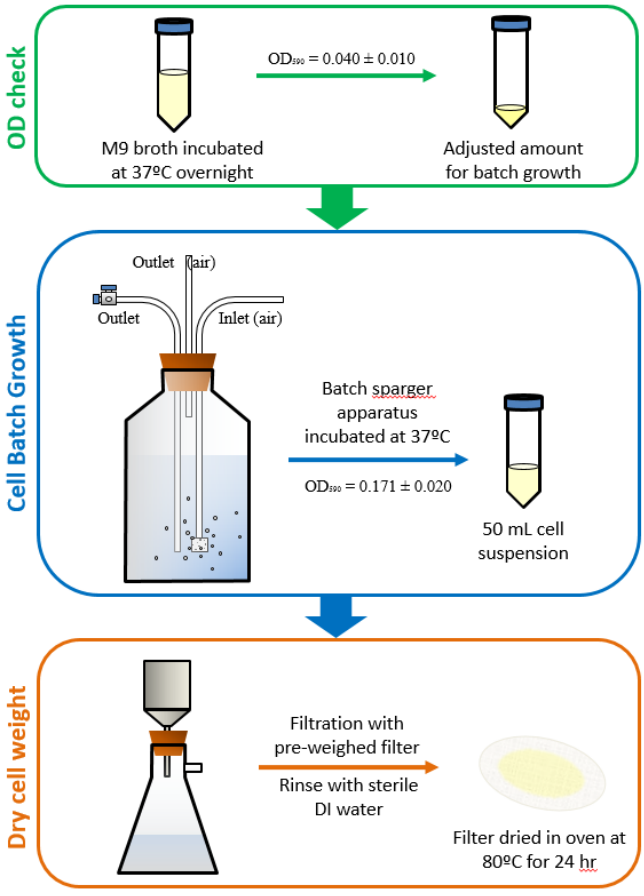

Figure 8: The flowchart of bacterial batch growth experiment for quantitative validation

GEM Prediction – Predictions were made using the draft GEMs on nutrient utilization (C, N, P and S substrates), biomass generation, and essential genes. By comparing the model predictions of nutrient utilization with the phenotype information obtained from the Biolog^{TM} PM experiments, potential exchange reactions (or transporter reactions) can be identified to extend the reaction list for the *in silico* stoichiometric metabolic networks. The level of agreement was determined to evaluate the accuracy of qualitative model predictions made by each GEM. For quantitative model prediction on bacterial growth, a conversion factor between biomass and OD_{600} was determined experimentally to estimate the initial biomass based on the initial OD_{600} so that value can be included for model prediction.

To improve the agreement between the predicted and experimentally determined growth rate (h^{-1}) and biomass yield, a scaling factor was calculated and introduced to adjust the biomass generation for each GEM as described in previous study [88]. The corresponding glucose uptake rate was also adjusted to match with the experiment value. Additionally, essential reactions were identified and compared among the strains by conducting FBA on GEMs to optimize the biomass generation rate while restricting each of the metabolic reactions one at a time to have zero flux. If the biomass prediction equals to zero, the restricted reaction is determined to be essential because biomass cannot be produced without this specific metabolic reaction [89]. These essential reactions represent potential targets of control for the organism.

IV. 3 **Results**

Chosen Strains – The 5 *Salmonella* strains in question display a variety of disease transmission routes according to their isolation sources and related outbreaks (Table 3). To evaluate the similarities and differences among these strains' genomes, a pangenome was constructed on KBase using the five strains' annotated genomes so the shared and unique protein coding genes across the entire input set of genomes can be compared. Each strain contains more than 90% commonly shared protein coding genes while 0.4% - 9.8% of their genes are unique (Table 6).

Table 6: The shared and unique protein coding genes among the chosen strains' genomes

Strain	Protein coding genes	Shared genes (%)	Unique genes (%)
LT2	4452	4389 (98.58%)	63 (1.42%)
UK1	4455	4437 (99.60%)	18 (0.40%)
CVM23701	4602	4151 (90.20%)	451 (9.80%)
ATCC35640	4175	3982 (95.38%)	193 (4.62%)
P125109	4502	4201 (93.31%)	301 (6.69%)

Mauve was used to visualize the evolutionary changes among the five strains by aligning the homologous genomic regions. As shown in Figure 9, each continuously colored genomic regain among all or partial genomes is called a locally collinear block (LCB). The homologous backbone sequences in the LCBs were not rearranged, and therefore those regions were conserved among all or within a subset of the 5 genomes. The colorless segments within an LCB, representing the non-conserved strain-specific sequence, can be identified more easily in the backbone view where genomic regions shared by all five genomes are colored in mauve (Figure 10). The genomes of str. LT2

and UK1 contain more conserved LCBs compared to the rest, while each genome contains unique genetic regions compared to a subset or the rest of the five genomes.

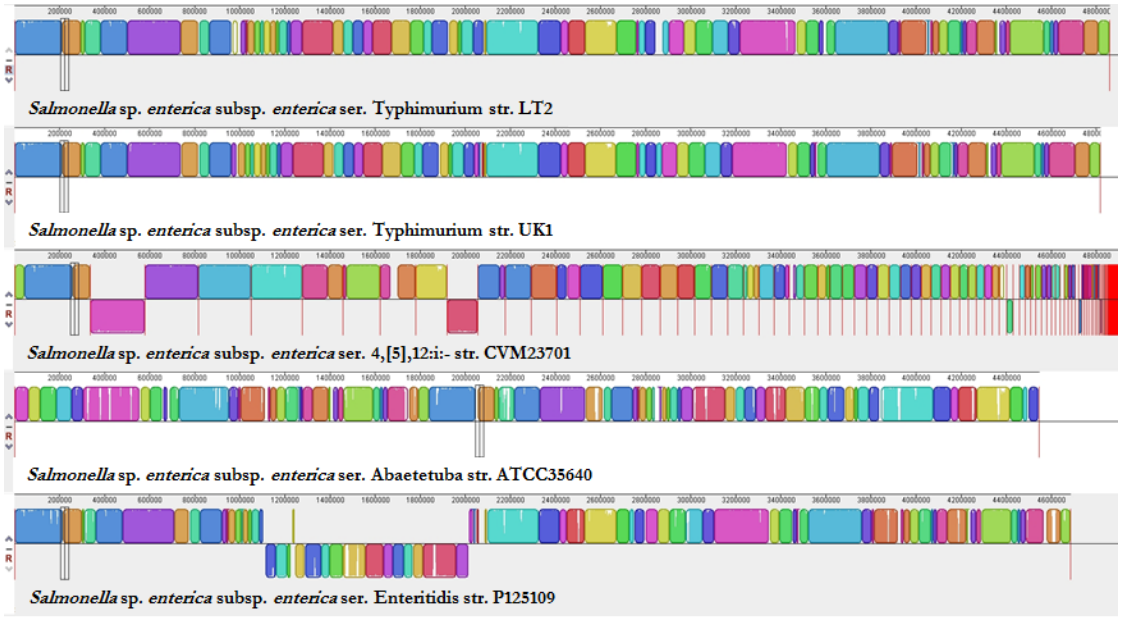

Figure 9: Locally collinear blocks identified among five *Salmonella* genomes through the Mauve rearrangement viewer

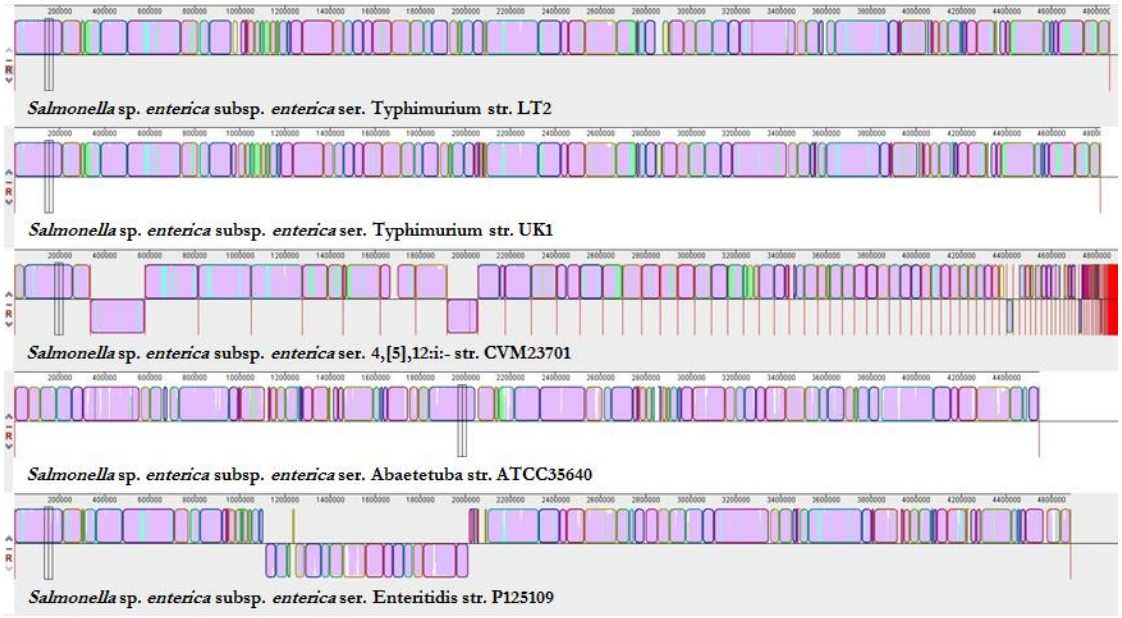

Figure 10: Mauve genome alignment in backbone view for strain-specific genes' identification

GEM Construction – The quality of a GEM depends of the amount of genes, metabolites, and metabolic reactions involved representing a complete metabolic network. All of the GEMs constructed in this study contain more than 1,000 annotated genes that encode over 1,600 metabolites catalyzing approximately 1,700 reactions as shown in Table 7. A biomass generation reaction (bio1) was formulated correspondingly, illustrating the metabolites required to form 1 g/DCW of bacterial cells.

Table 7: The number of genes, metabolites and reactions contained by each GEM

Serovar	Strain	# Genes	# Metabolites	# Reactions (Draft GEMs)	# Reactions (Refined GEMs)
Typhimurium	LT2	1,153	1,706	1,695	1,749
Typhimurium	UK1	1,144	1,695	1,684	1,749
4,[5],12:i:-	CVM23701	1,059	1,690	1,645	1,710
Abaetetuba	ATCC 35640	1,103	1,685	1,670	1,726
Enteritidis	P125109	1,152	1,691	1,676	1,741

By comparison, 1,613 reactions and 1,535 metabolites were found commonly shared among the draft GEMs. After qualitative model validation, the manual curation added about 60 exchange reactions to each of the GEMs, and the number of these conserved reactions was increased to 1,680 consequently. The number of unique reactions in stoichiometric matrix that belong to a single GEM compared to others was listed in Table 8 (for details see Supplement file 1. Detailed GEM information). Additionally, the GEM of str. CVM23701 contains four unique reactions compared to the rest of the GEMs (rxn00125_c0: S-Adenosyl-L-methionine hydrolase_c0, rxn02304_c0: protoporphyrinogen-IX:oxygen oxidoreductase_c0, rxn10954_c0: Fatty acid biosynthesis

(n-C18:0)_c0, and rxn00399_c0: L-Arginine,NADPH:oxygen oxidoreductase (nitric-oxide-forming)_c0.)

Table 8: The number of unique reactions belonging to one GEM (top row) compared to others

	LT2	UK1	ATCC35640	P125109	CVM23701
LT2		1	1	3	6
UK1	12		12	2	5
ATCC35640	26	26		20	30
P125109	22	10	14		12
CVM23701	56	44	55	43	

Qualitative Validation – The Biology^{TM} Phenotype MicroArray (PM) experiments were conducted to determine the three chosen strains metabolic capabilities on 95 carbon substrates (PM1), 95 nitrogen substrates (PM3), 59 phosphorus substrates (PM4), and 35 sulfur substrates (PM4). Assuming the strains can catabolize all these nutrient substrates and exchange them between the cytosol and extracellular space for metabolism, 21 carbon substrate exchange reactions, 20 nitrogen substrate exchange reactions, 11 phosphorus substrate exchange reactions, and 1 sulfur substrate exchange reactions were added into all of the GEMs' reaction list after surveying the metabolites present in their stoichiometric matrixes. Besides, 7 carbon substrate exchange reactions and 7 nitrogen substrate exchange reactions found only in one or a subset of the GEMs were added to the rest of the GEMs. As a result, each updated GEM contains 75 out of the 95 carbon substrates, 69 out of the 95 nitrogen sources, 33 out of the 59 phosphorus substrates, and 13 out of the 35 sulfur substrates tested by the Biolog^{TM} PMs. These nutrient substrates

then were used for the comparison between experimental results and model predictions on nutrient utilization to determine the accuracy of qualitative model predictions.

The level of agreement between *in silico* and *in vitro* data for nutrient utilization on carbon (Figure 11), nitrogen (Figure 12), phosphorus and sulfur (Figure 13) were determined using qualitative data (Y = bacterial growth, N = no bacterial growth) and the determined percentages of agreement were concluded in Table 9. The model prediction agrees with the experiment result if the model predicted growth (i.e. biomass production) by metabolizing a particular nutrient substrate and experiment also showed bacterial growth, or both model prediction and experiment result showed no bacterial growth. False positives occur when model predicted bacterial growth while experiment showed none, and false negative means model predicted no bacterial growth while experiment showed the opposite (for details see Supplement file 2. Bacterial phenotype information).

Table 9: Percentages of agreement between model predictions and experiment results for the GEMs of three *Salmonella* strains

Nutrient utilization	aerobic/ anaerobic	LT2	UK1	ATCC35640
Carbon	aerobic	100%	96%	91%
	anaerobic	96%	96%	95%
Nitrogen	aerobic	62%	61%	61%
Phosphorus	aerobic	61%	61%	61%
Sulfur	aerobic	8%	8%	8%

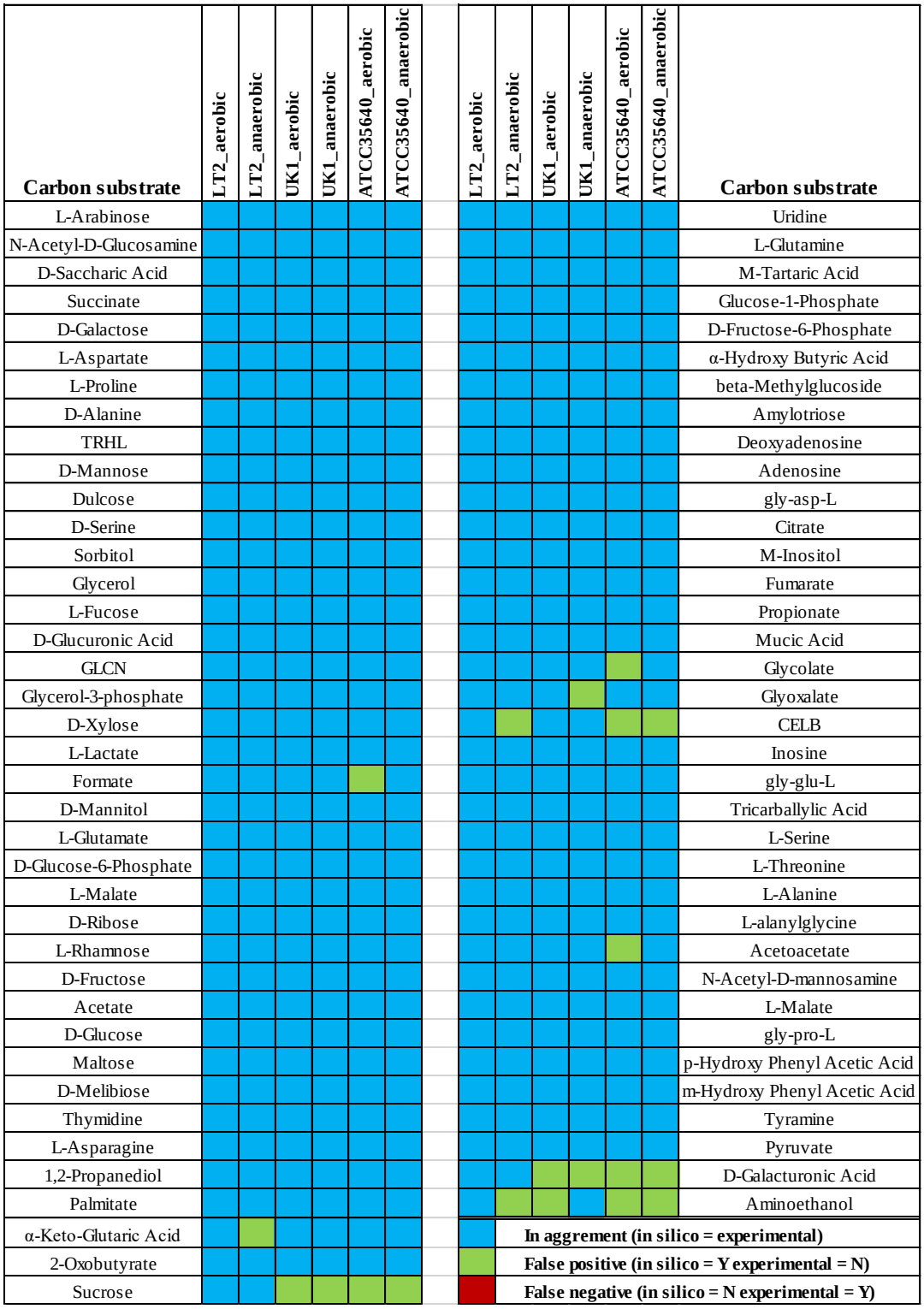

Figure 11: Comparison between model predictions and experiment results on 75 carbon substrates' utilization for the three strains under aerobic and anaerobic conditions

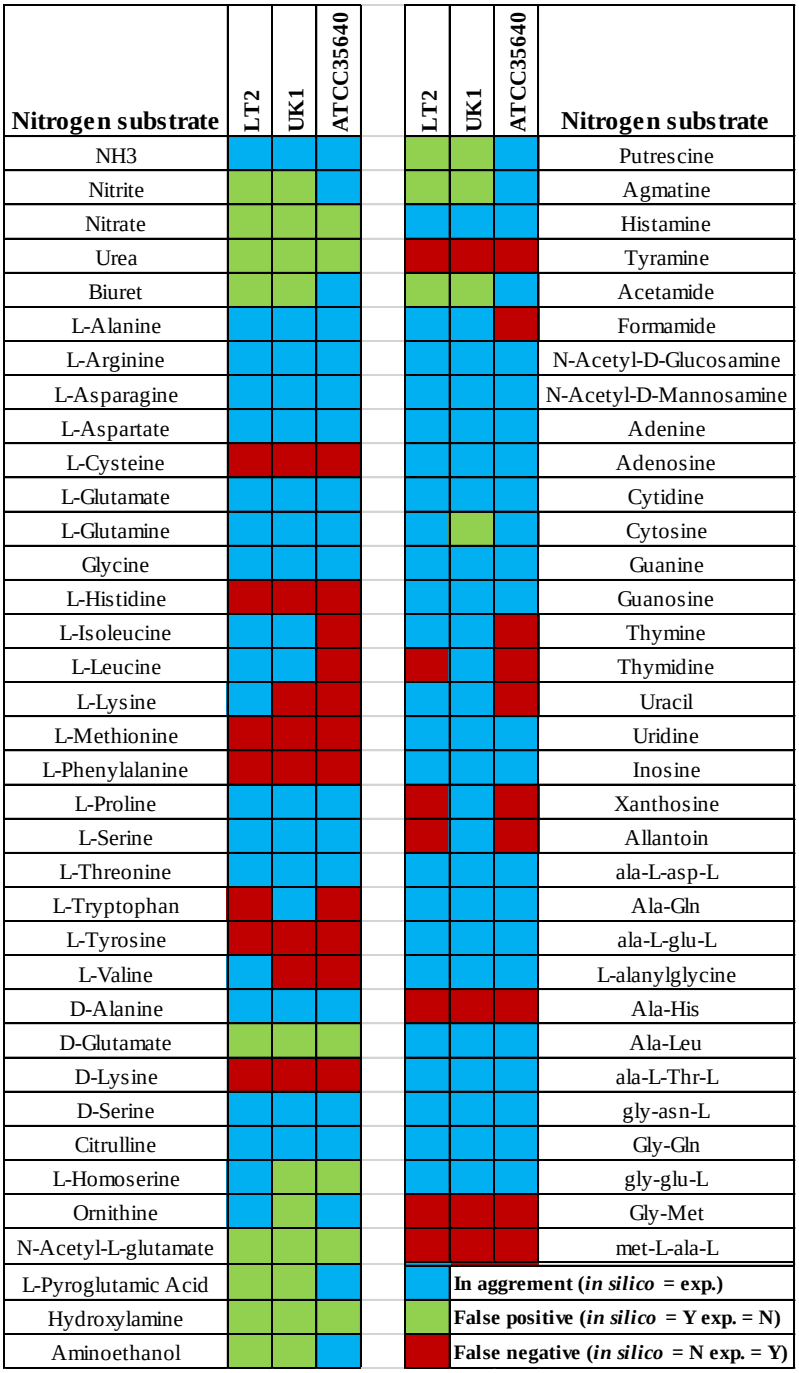

Figure 12: Comparison between model predictions and experiment results on 69 nitrogen substrates' utilization for the three strains under aerobic condition

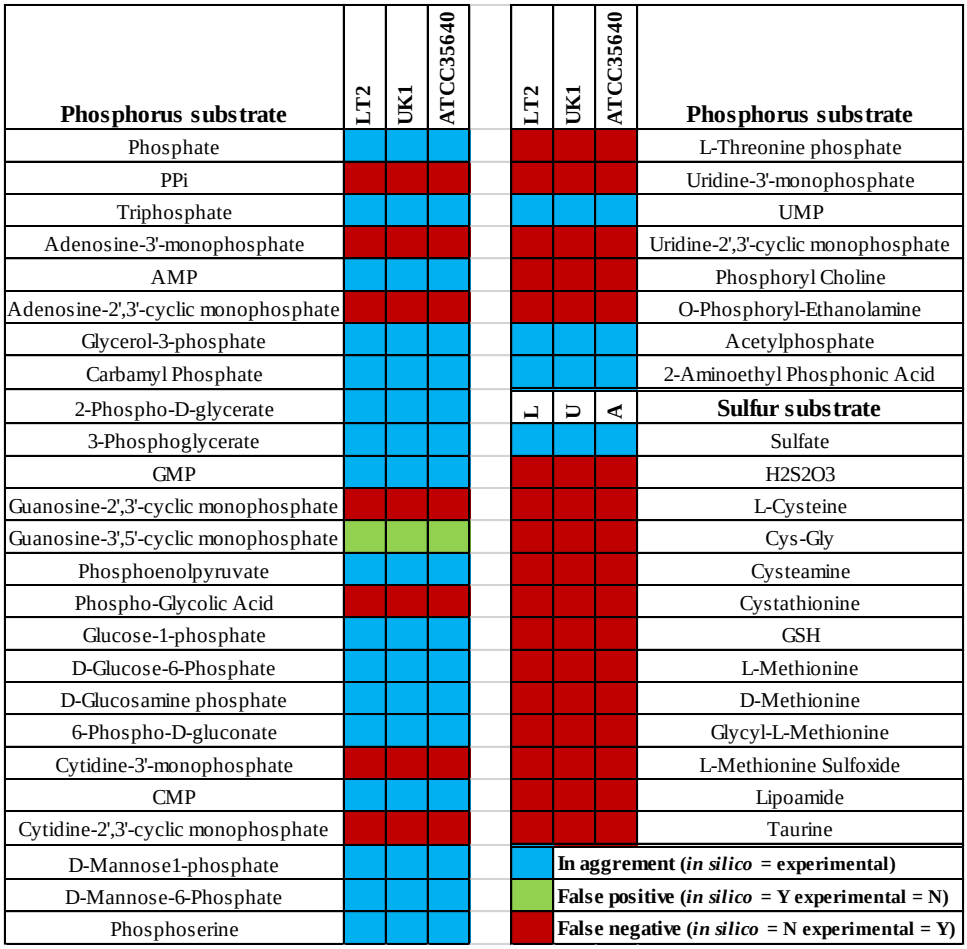

Figure 13: Comparison between model predictions and experiment results on 33 phosphorous and 13 sulfur substrates' utilization for the three strains under aerobic condition

Quantitative Validation –The M9 minimal medium used for the bacterial batch growth experiment contains defined chemicals, Na_2HPO_4 (6 g/L), KH_2PO_4 (3 g/L), NaCl (0.5 g/L), NH_4Cl (1 g/L), $MgSO_4$ (2 mM), $CaCl_2$ (0.1 mM) and a single carbon substrate (glucose). Consistent growth was observed as shown in Figure 14 for the three *Salmonella* strains growing on this medium, and therefore the M9 minimal medium was acceptable for bacterial batch growth experiments to refine FBA parameters including initial biomass and biomass generation rate for quantitative model validation.

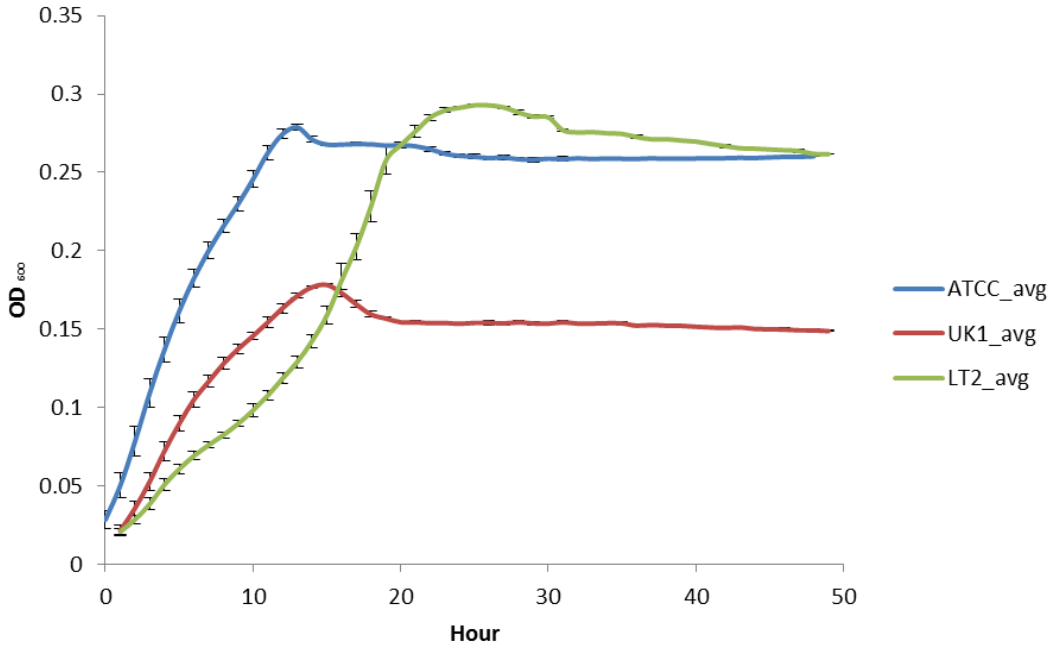

Figure 14: Bacterial growth curve generated for the three *Salmonella* strains incubated with M9 minimal medium at 37°C for 48 hours

The linear correlation between each strain's biomass (DCW) and corresponding OD_{600} measurements was significant ($R^2 > 0.95$), thus the correlation coefficient can be defined as the conversion factor to estimate the value of the initial biomass from the initial OD_{600} value for each GEM (Table 10).

Table 10: Experimentally determined OD600 – biomass conversion factor and initial biomass

Strain	Aerobic/ Anaerobic	OD_{600} to Biomass (gDCW/L) ± SD	R^2	Initial OD_{600}	Initial Biomass (g/L)
ATCC 35640	aerobic	0.609 ± 0.012	0.995808	0.033	0.020
	anaerobic	0.578 ± 0.017	0.991647	0.041	0.024
UK 1	aerobic	0.567 ± 0.012	0.995478	0.038	0.022
	anaerobic	0.545 ± 0.013	0.994079	0.037	0.020
LT 2	aerobic	0.596 ± 0.010	0.997324	0.038	0.023
	anaerobic	0.476 ± 0.034	0.952319	0.039	0.018

Plots were generated to visually present the mathematical correlation between the biomass and OD_{600} measurements (Figure 15) (for details see Supplement file 3. Bacterial growth information).

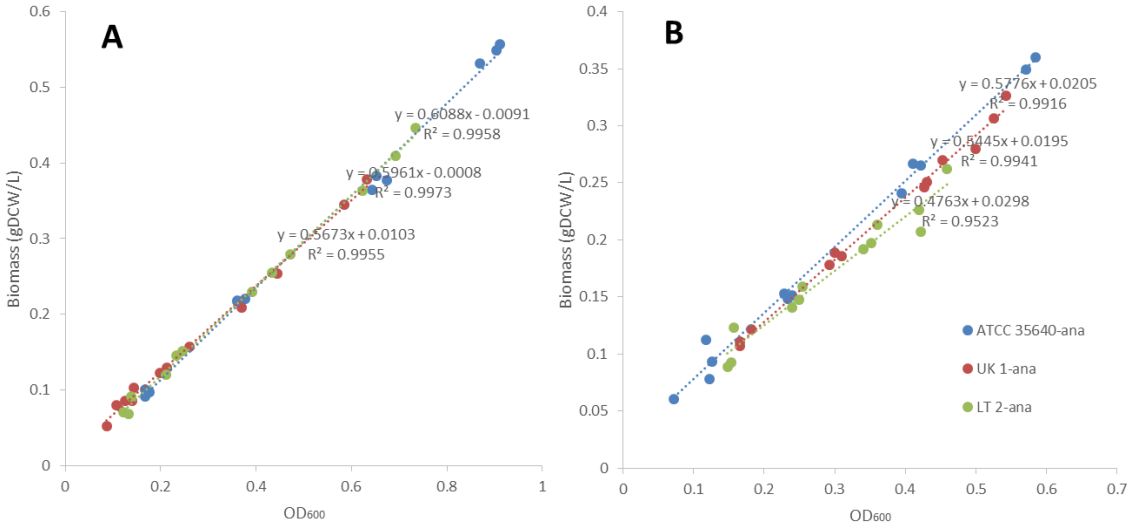

Figure 15: Biomass v.s. OD600 for batch growth experiment under aerobic (A) and anaerobic (B) conditions

The experimental growth rate (h^{-1}), glucose utilization rate (mmol• $h^{-1}L^{-1}$) and glucose uptake rate (g DCW/ g glucose) were also determined through linear regression, and a scalar factor was defined for each GEM to adjust the model predictions. The correlation between the experimental and predicted growth rates was not significant ($0.01 < R^2 < 0.75$) before the scalar factors were introduced while it became significant after the determination of these scalar factors ($R^2 > 0.96$). The same change occurred to the predictions on glucose utilization rate (mmol• $h^{-1}L^{-1}$) as shown in Table 11.

Table 11: Comparison of experimental and model predicted growth rates and glucose utilization rates

	Growth rate (h^{-1})			Glucose utilization rate (mmol● $h^{-1}L^{-1}$)		
	Experiment	Prediction initial (R^2)	Prediction final (R^2)	Experiment	Prediction initial (R^2)	Prediction final (R^2)
LT2 _aerobic	0.56	0.37 (0.63)	0.48 (1.00)	2.18	4.36 (0.42)	2.31 (1.00)
UK1 _aerobic	0.42	0.37 (0.60)	0.42 (1.00)	2.97	4.46 (0.43)	1.63 (1.00)
ATCC35640 _aerobic	0.57	0.37 (0.75)	0.55 (0.97)	4.64	4.66 (0.53)	2.78 (0.97)
LT2 _anaerobic	0.28	0.00 (0.01)	0.44 (0.96)	4.27	0.00 (0.00)	2.53 (0.92)
UK1 _anaerobic	0.33	0.00 (0.02)	0.50 (0.96)	5.02	0.00 (0.00)	3.74 (0.91)
ATCC35640 _anaerobic	0.47	0.50 (0.69)	0.51 (0.99)	4.86	4.47 (0.51)	2.86 (0.98)

The glucose uptake rate was determined as the amount of DCW generated (g) per glucose consumed (g). With p values approximate to 1, the correlation between these two variables is significant for the experimental results and the predictions made with or without the refined scaling factors. However, the initial predicted values were about 10 times larger than the experimental results, while they became closer to the *in vitro* values after the scaling factors were introduced (Table 12).

Table 12: Comparison of experimental and model predicted glucose uptake rates

	LT2		UK1		ATCC35640	
	Aerobic	Anaerobic	Aerobic	Anaerobic	Aerobic	Anaerobic
Experimental result (R^2)	0.19 (0.901)	0.06 (1.000)	0.12 (1.000)	0.07 (1.000)	0.18 (1.000)	0.09 (1.000)
Prediction initial (R^2)	0.79 (0.999)	0.70 (0.998)	0.76 (0.999)	0.78 (0.996)	0.68 (1.000)	0.95 (0.990)
Prediction final (R^2)	0.27 (0.999)	0.14 (0.997)	0.23 (0.999)	0.15 (1.000)	0.30 (1.000)	0.16 (0.995)

GEM Prediction – The GEMs were first used to make qualitative predictions on nutrient utilization. Then they were performed to predict essential reactions on simulated M9 (default) and two food items. By comparison, these five GEMs share approximate 331 essential reactions, including the exchange reactions for phosphate, NH_3, Mn^{2+}, Zn^{2+}, sulfate, Cu^{2+}, Ca^{2+}, Cl^-, Co^{2+}, K^+, and Mg^{2+}. The GEM for CVM23701 involves Fe^{2+} and Fe^{2+}'s exchange reactions as essential reactions specifically. Differences in essential reactions belong to each GEM were found under different nutrient conditions as shown in Table 13 (for details see Supplement file 4. Essential reactions).

Table 13: Essential reactions (highlighted in yellow) present for model predictions under three different conditions

	Default										Egg white (raw)										Chicken breast (raw, breaded)									
	LT2		UK1		ATCC35640		P125209		CVM23701		LT2		UK1		ATCC35640		P125209		CVM23701		LT2		UK1		ATCC35640		P125209		CVM23701	
Reaction	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana	aer	ana
EXpd105150																														
EXpd105160																														
rxn000150																														
rxn001270																														
rxn001790																														
rxn006530																														
rxn008020																														
rxn009120																														
rxn009150																														
rxn014050																														
rxn014060																														
rxn014340																														
rxn023730																														
rxn023740																														
rxn024840																														
rxn039780																														
rxn051950																														
rxn053470																														
rxn053490																														
rxn055550																														
rxn085510																														
rxn088080																														
rxn088090																														
rxn094490																														
rxn102130																														
rxn125120																														
rxn099970																														

After the parameters including initial bacterial biomass and scalar factors were experimentally defined and set for FBA, the GEMs were able to make quantitative perditions for bacterial growth. The resulting biomass generation curves were plotted along with the experimentally determined curves for aerobic and anaerobic bacterial growth and glucose utilization (Figure 16). The model predictions correlated well with experimental data in the log phase, and the overall Pearson correlation coefficients are good for the three GEMs (> 0.94).

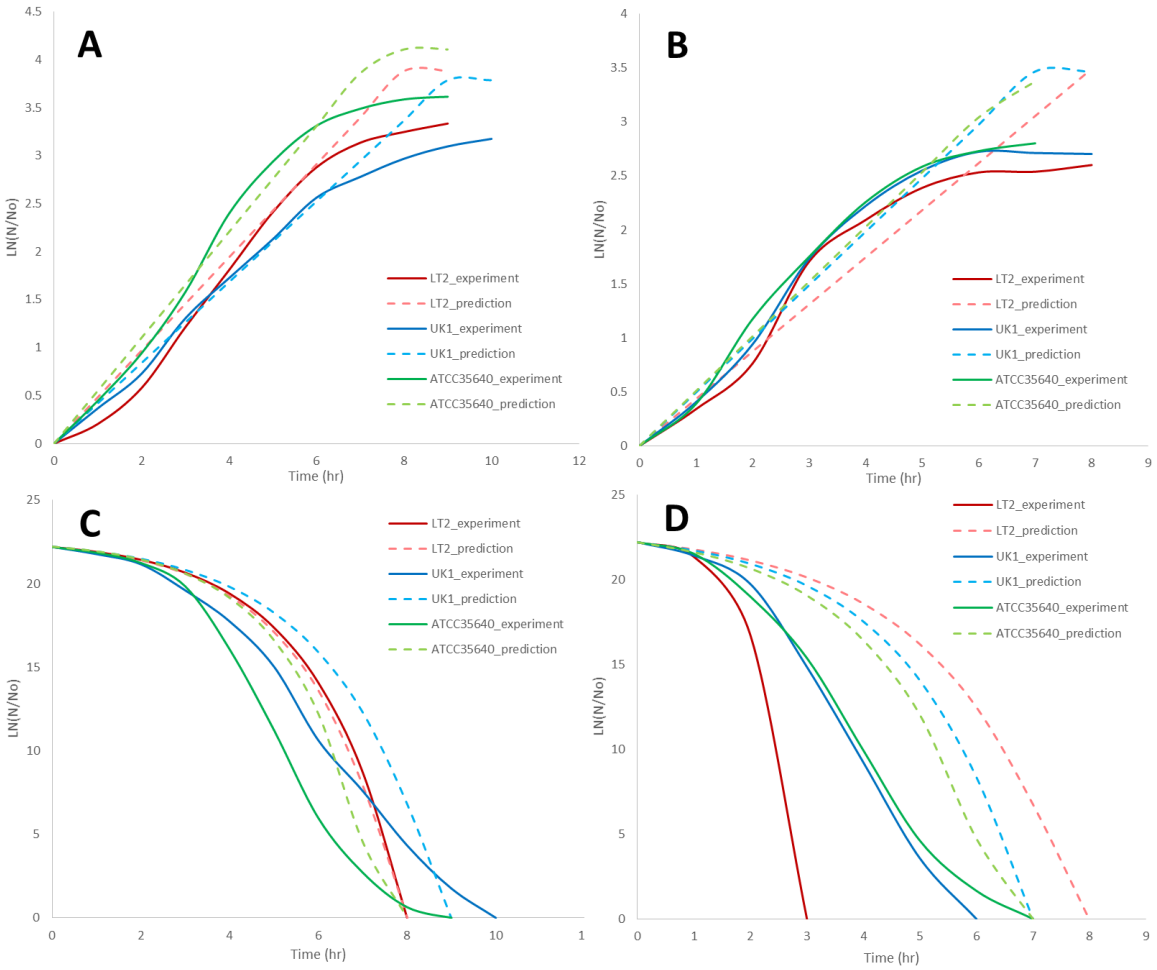

Figure 16: Comparison between computationally predicted and experimentally determined bacterial batch growth under aerobic (A) and anaerobic (B) conditions, and glucose utilization under aerobic (C) and anaerobic (D) conditions

IV. 4 **Discussion**

Chosen strains – The genome of a bacterium contains essential genes that encode a number of basic metabolic reactions functioning to assist cell survival. During bacterial evolution, the microorganism may acquire or lose specific metabolic capabilities due to gene mutation, rearrangement, and especially horizontal transfer of certain genomic regains aside from those core genes, which shapes the bacterium's viability and adaptation encountering certain environmental stresses.[78, 90] The five *Salmonella* strains studied in this work display different disease transmission routes and host preference as shown in Table 4, and the dissimilarity in their gene contents may be the answer to what has caused those differences.

Genes encoding proteins are often linked to metabolic functions, and KBase is able to compare protein coding genes based on a pangenome constructed using all of the five strains' genomes. The comparison result showed most of those genes are conserved among the five *Salmonella* strains (> 90%), and the difference in these strains' metabolic capabilities is caused by a small portion of their protein coding genes (Table 6). The str. ATCC35640 isolated from the natural environment contains the least protein coding genes (4175) compared to the other disease-causing *Salmonella* strains. It probably has not acquired the essential metabolic genes during evolution, so the strain is not capable to live and cause infection in niches in animal hosts. The str. CVM23701 from the serovar 4,5,12:i:- possesses the most protein coding genes (4602) and the most unconserved genes (9.8%), which may relate to its acquiring on genes that are responsible for unique metabolic reactions and multidrug resistance genes during bacterial evolution.

The Mauve genome alignment directly visualized the homologous genes conserved among the 5 chosen strains' chromosomal genomes (Figure 9). Although the LCBs are internally free from genetic rearrangement, many of them were found reordered on one genome compared to the other due to genomic recombination possibly during bacterial evolution. Belonging to the same serovar, str. UK1 and LT2 revealed high collinearity on their genomes with more conserved LCBs compared to the rest. With the mauve backbone view, non-conserved genes can be easily found that are present only in a single or a subset of genomes (Figure 10). Each strain's genome has undergone unique genome rearrangements compared to others, and these differences may be attributable to the acquisition of unique metabolic capabilities or even virulence factors that benefit the strain under certain environment.

GEM Construction – The network reconstruction accounts for all of the biosynthesis pathways required for bacterial growth, including central metabolism, amino acid and nucleotide biosynthesis, and nutrient compound utilization; therefore, the genes, metabolites, and reactions possessed by an *in silico* metabolic network are the important characteristics that shaped each of the models. All of the GEMs constructed in this study contain more reactions compared to the three previously established *Salmonella* models as shown in Table 14. Both previously published GEMs iMA945 and iRR1083 were constructed in 2009 based on well-established *E. coli* GEMs (iAF1260 and iJR904), and the model STM_v1.0 was formed one year later combining the previously established two *Salmonella* GEMs [43, 47, 48]. All of the current published bacteria GEMs including above three *Salmonella* GEMs can be found at http://darwin.di.uminho.pt/models.

Table 14: The number of genes, metabolites, and reactions contained in the GEMs constructed in this study and the previously published GEMs

GEM	# Genes	# Metabolites	# Reactions
S. LT2 (in this study)	1,153	1,706	1,749
S. UK1 (in this study)	1,144	1,695	1,749
S. CVM23701 (in this study)	1,059	1,690	1,710
S. ATCC 35640 (in this study)	1,103	1,685	1,726
S. P125109 (in this study)	1,152	1,691	1,741
S. LT2 (iMA945, 2009)	945	1,964	1,036
S. LT2 (iRR1083, 2009)	1,083	1,087	973
S. LT2 (STM_v1.0, 2011)	1,270	2,201	1,119

Although the majority of the 5 chosen strains' chromosomal genome contents are similar to each other, the metabolic networks reconstructed *in silico* through KBase are able to distinguish the differences in metabolic reactions and metabolites encoded by chromosomal genes for each chosen strain (Table 8). The genome contents of strain LT2 and UK1 that under the same serovar are more similar to each other compared to others, while str. LT2's GEM contains about 10 more transporter reactions than the GEM of UK1.

Specifically, the GEM of str. P125109 contains 1 reaction that cannot be found in other GEMs (L-arabinose transport via a ABC transport system), while str. CVM23701's GEM contains 4 such reactions (S-Adenosyl-L-methionine hydrolase, protoporphyrinogen-IX:oxygen oxidoreductase, Fatty acid biosynthesis (n-C18:0), and L-Arginine,NADPH:oxygen oxidoreductase (nitric-oxide-forming)). The unique reaction present in str. P125109' GEM was used to expand the reaction list for all the other

GEMs, considering it is an exchange reaction of L-arabinose and those *Salmonella* strains should all be able to transport and metabolize L-arabinose based on the Biolog^{TM} PM result.

The 4 reactions that are unique to the str. CVM 23701's GEM were kept unique because they create unique metabolic capabilities for that particular strain. For example, the GEM of str. CVM23701 possess a unique protoporphyrinogenIX oxidase reaction that requires the use of oxygen to produce protoporphyrin (rxn02304: (3) O2_c0 + (2) ProtoporphyrinogenIX_c0 -> (6) H2O_c0 + (2) Protoporphyrin_c0), while in other strains' GEMs another protoporphyrinogen oxidase exists instead that can also produce the same metabolite even under anaerobic condition (rxn09180: (3) Fumarate_c0 + (1) ProtoporphyrinogenIX_c0 <-> (3) Succinate_c0 + (1) Protoporphyrin_c0).

The L-Arginine production reaction ((1) NADPH_c0 + (3) O2_c0 + (2) L-Arginine_c0 <- (2) H2O_c0 + (1) NADP_c0[c0] + (1) H+_c0 + (2) Citrulline_c0 + (2) NO_c0) included in this strain's GEM is also found unique, and this reaction is responsible for citrulline metabolism. The utilization of citrulline may contribute to the acid tolerance of bacteria, and its function in soft tissue infection has been approved on a murine model for *Streptococcus pyogenes* infection [91]. Considering swine can be fed with a wide range of food, including fermented feed, it is reasonable that the swine related str. CVM 23701 possesses this unique metabolic reaction along with other unique reactions.

Qualitative prediction – Phenotype data were collected from the Biolog PM experiment to compare with *in silico* predictions on chosen strains' nutrient utilization capabilities and to determine the degree to which the model predictions agree with experiment results for qualitative model validation (Table 9). Significantly, the GEMs constructed in this work can be used to make more accurate predictions on carbon substrates utilization compared to all other previously published GEMs (% agreement > 90%) (Figure 17).

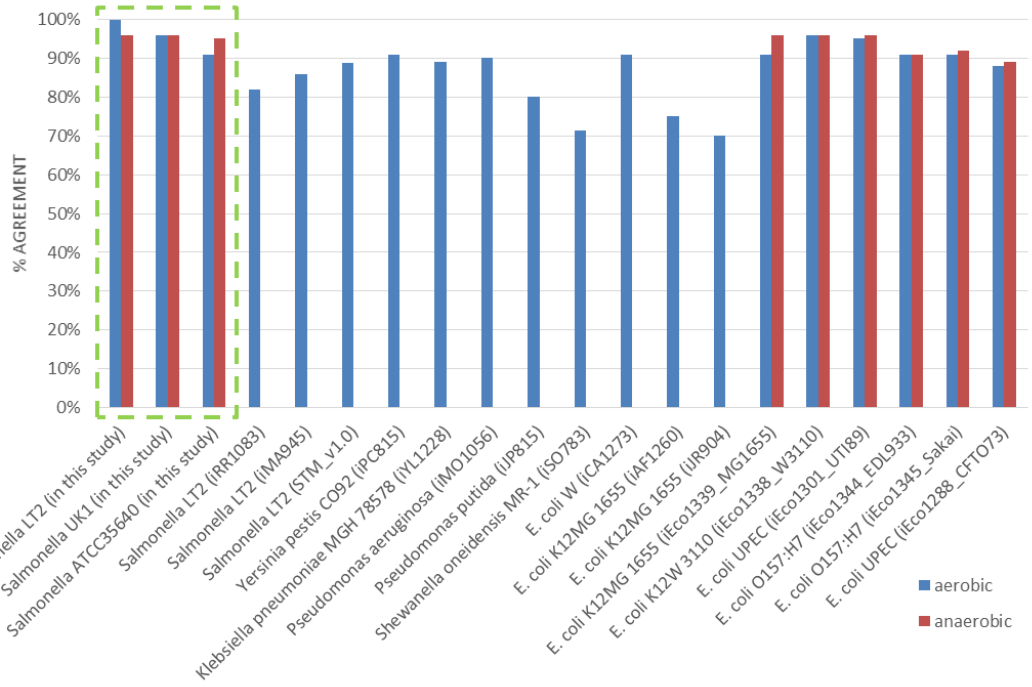

Figure 17: The percentages of agreement between model predictions and experimental data on carbon sources utilization for 17 previously published GEMs and the three GEMs constructed in this study

Compared to the well-studied carbon sources related bacterial metabolism, the mechanism of other nutrient sources' metabolism was not elucidated enough yet. The GEMs constructed in this work only identified 13 sulfur substrates' metabolic reactions

retrieved from the public database to compare with experimental data, and the comparison results on nitrogen, phosphorus and sulfur substrates' metabolism indicated an area for improvement for these models (Figure 18). If more studies can be done on such nutrients' metabolism to keep the database updated consistently, more reactions can be used to improve the GEMs by completing their reconstructed metabolic network.

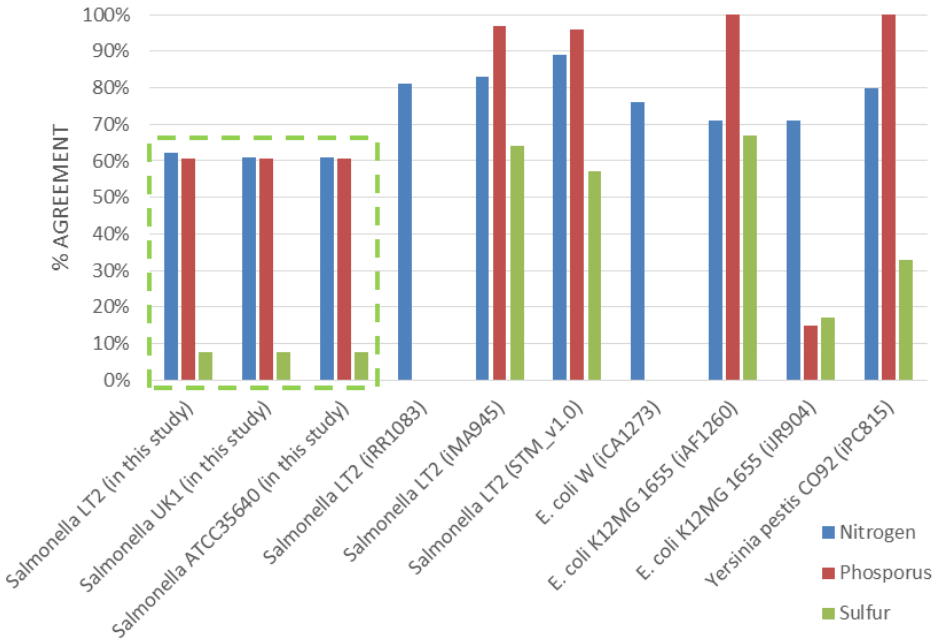

Figure 18: The percentages of agreement between *in silico* and experimental data on N, P, and S sources utilization perditions for the GEMs generated in this study and previously published GEMs

Biolog PMs use tetrazolium redox dye to colorimetrically measure cell respiration, thus identifying cellular phenotypes. Not all the metabolic pathways lead to cell growth, and therefore measuring cell respiration instead of cell growth makes this method more sensitive in cellular pathways detection.[92] However, the objective function used in this study is biomass generation, not cell respiration. If the bacteria can respire (experimental = Y) but are not capable to grow (*in silico* = N) upon certain nutrients, the prediction will

show a false negative result and this may be another reason to why there were many false negative results generated especially for phosphorus and sulfur sources utilization.

With the qualitatively validated GEMs, essential genes related essential reactions were predicted for all 5 strains living aerobically and anaerobically on computationally simulated M9 medium, egg white and chicken breast living conditions. Figure 13 revealed the differences in the identified essential reactions among the five different strains. The same strain's essential reactions also change based on the strain's living environment. For example, the reaction rxn014340 is essential when str. UK1 lives in the minimal medium, while it is not considered essential when the strain lives on the other two food items. This reaction is related to L-aspartate conversion, one of the metabolic reactions for bacterial survival, and L-aspartate is contained in both egg white and chicken breast but not in the M9 medium. Therefore, it is possible that str. UK1 considers this reaction an essential reaction only when its living environment lacks L-aspartate. The GEMs constructed in this work can be used to make such predictions to better understand bacterial metabolic capabilities under different conditions.

Quantitative prediction – The M9 medium was chosen as the defined nutrient medium for *Salmonella* batch growth experiment in this study, considering it has been successfully used in many *E. coli* studies. In order to utilize the data from batch growth experiments to quantitatively validate the GEMs, the media used to perform the batch growth experiment must be chemically defined so that the nutrient amounts can be accurately translated in GEMs for *in silico* growth simulation. M9 contains finite amount of defined chemicals that can be quantitatively used for growth simulations of each of the

models for FBA, and the three *Salmonella* strains are able to utilize those limited nutrient compounds to live and proliferate. Therefore, this medium is verified to be useful for quantitative GEM study on *Salmonella*.

For the bacterial batch growth experiment, dry cell weight was determined and a significant linear correlation was found between each strain's biomass (DCW) and corresponding OD_{600} readings ($R^2 > 0.95$) (Figure 15). The initial biomass and scalar factors were then estimated for each strain respectively under aerobic and anaerobic conditions, and by incorporating these parameters the GEMs can be used to make predictions that display high correlation with the experimental data (Table 10). Figure 16 (A) visually indicated that the GEMs constructed in this work can be used to make accurate predictions on bacterial growth during the log phase under aerobic condition based on the comparison between the *in silico* and experimental defined biomass generation curves. Compared to the other strains, str. LT2's GEM constructed in this study can be used to make accurate predictions on bacterial glucose uptake under aerobic condition (Figure 16 C).

The inaccuracy of model predictions may be caused by the unpredictable difficulties that bacteria may encounter during their growth, which make the model predictions become unrealistic after the log phase period compared to the experimentally determined bacterial growth curves. For example, it may be hard for one bacterial cell crowded by other cells to move and find nutrients for its survival and proliferation if there are too many of them present in a limited area. To improve the accuracy of

quantitative model predictions, these environmental factors should also be considered for GEMs construction.

Overall, the GEMs constructed in this study can be used to make accurate predictions on bacterial phenotype behavior (nutrient utilization) and decant predictions on bacterial growth. KBase allows computational systems biology analysis through the RAST server, and the continuously updating of the subsystems library awards RAST's accuracy, consistency, and completeness compared to other tools [93]. This work has demonstrated the usefulness of the semi-automated platform KBase to generate *Salmonella* GEMs with great analysis potential, and thus will be useful to conduct more large scale analysis of >100's of genomes from a particular *Salmonella* serovar to ascertain genome-level differentiating metabolic capabilities. Future studies using GEMs will allow a new unprecedented amount of elucidation of bacterial evolution and host-microbe interactions in this era where 1,000,000's of genomes of microorganisms are currently being sequenced through next generation sequencing technologies.

Reference List

1. **The 10 Deadliest Outbreaks in U.S. History — Revisited** [http://www.foodsafetynews.com/2012/04/the-ten-deadliest-outbreaks-in-history-revisited]
2. Brooks J: **The sad and tragic life of Typhoid Mary**. *CMAJ: Canadian Medical Association Journal* 1996, **154**(6):915.
3. Tauxe RV: **Emerging foodborne diseases: an evolving public health challenge**. *Emerging infectious diseases* 1997, **3**(4):425.
4. **A History of Salmonella Lawsuits** [http://www.foodpoisonjournal.com/foodborne-illness-outbreaks/a-history-of-salmonella-lawsuits]
5. Jones TF, Scallan E, Angulo FJ: **FoodNet: overview of a decade of achievement**. *Foodborne Pathogens and Disease* 2007, **4**(1):60-66.
6. Control CfD, Prevention: **Vital signs: incidence and trends of infection with pathogens transmitted commonly through food--foodborne diseases active surveillance network, 10 US sites, 1996-2010**. *MMWR Morbidity and mortality weekly report* 2011, **60**(22):749.
7. Török TJ, Tauxe RV, Wise RP, Livengood JR, Sokolow R, Mauvais S, Birkness KA, Skeels MR, Horan JM, Foster LR: **A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars**. *Jama* 1997, **278**(5):389-395.
8. Tucker JB: **Historical trends related to bioterrorism: An empirical analysis**. *Emerging infectious diseases* 1999, **5**(4):498.
9. Dominguez SA, Schaffner DW: **Survival of salmonella in processed chicken products during frozen storage**. *Journal of food protection* 2009, **72**(10):2088-2092.
10. **SALMONELLA** [http://www.foodsafetywatch.org/factsheets/salmonella/]
11. Administration FaD: **Bad Bug Book, Foodborne Pathogenic Microorganisms and Natural Toxins**, Second edn; 2012.
12. Gal-Mor O, Boyle EC, Grassl GA: **Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ**. *Frontiers in microbiology* 2014, **5**:391.
13. Gunn JS: **Salmonella Host–Pathogen Interactions: A Special Topic**. *Frontiers in microbiology* 2011, **2**.
14. Popoff MY, Bockemühl J, Gheesling LL: **Supplement 2001 (no. 45) to the Kauffmann–White scheme**. *Research in Microbiology* 2003, **154**(3):173-174.
15. Popoff MY: **Antigenic Formulas of the Salmonella Serovars, 8th rendition**. *WHO Collaborating Centre for* 2001.
16. Nair S, Wain J, Connell S, de Pinna E, Peters T: **Salmonella enterica subspecies II infections in England and Wales–the use of multilocus sequence typing to assist serovar identification**. *Journal of medical microbiology* 2014, **63**(6):831-834.
17. Bale J, Meunier D, Weill F-X, dePinna E, Peters T, Nair S: **Characterization of new Salmonella serovars by whole-genome sequencing and traditional typing techniques**. *Journal of Medical Microbiology* 2016, **65**(10):1074-1078.
18. Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, Tewolde R, Schaefer U, Jenkins C, Dallman TJ: **Identification of Salmonella for public health surveillance using whole genome sequencing**. *PeerJ* 2016, **4**:e1752.
19. Hennessy TW, Hedberg CW, Slutsker L, White KE, Besser-Wiek JM, Moen ME, Feldman J, Coleman WW, Edmonson LM, MacDonald KL *et al*: **A National**

Outbreak of Salmonella enteritidis Infections from Ice Cream. *New England Journal of Medicine* 1996, **334**(20):1281-1286.
20. O'Neil J: **Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations 2014**. In.; 2014.
21. Verma S, Singh S: **Current and future status of herbal medicines**. *Veterinary world* 2008, **1**(11):347-350.
22. Kinch MS, Patridge E, Plummer M, Hoyer D: **An analysis of FDA-approved drugs for infectious disease: antibacterial agents**. *Drug discovery today* 2014, **19**(9):1283-1287.
23. Langdon A, Crook N, Dantas G: **The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation**. *Genome medicine* 2016, **8**(1):39.
24. Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, Finlay BB: **Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection**. *Infection and immunity* 2008, **76**(10):4726-4736.
25. McDevitt D, Rosenberg M: **Exploiting genomics to discover new antibiotics**. *TRENDS in Microbiology* 2001, **9**(12):611-617.
26. NARMS: **2014 Human Isolates Surveillance Report**. In. www.cdc.gov; 2014.
27. Rodrigue DC, Tauxe RV, Rowe B: **International increase in Salmonella enteritidis: a new pandemic?** *Epidemiology and infection* 1990, **105**(01):21-27.
28. Ghazaey S, Mirmomeni MH: **Microbial-resistant Salmonella enteritidis isolated from poultry samples**. *Reports of biochemistry & molecular biology* 2012, **1**(1):9.
29. Gantois I, Ducatelle R, Pasmans F, Haesebrouck F, Gast R, Humphrey TJ, Van Immerseel F: **Mechanisms of egg contamination by Salmonella Enteritidis**. *FEMS microbiology reviews* 2009, **33**(4):718-738.
30. Luo Y, Kong Q, Yang J, Golden G, Wanda S-Y, Jensen RV, Ernst PB, Curtiss R: **Complete genome sequence of the universal killer Salmonella enterica Serovar Typhimurium UK-1 (ATCC 68169)**. *Journal of bacteriology* 2011, **193**(15):4035-4036.
31. Lilleengen K: **Typing of Salmonella dublin and Salmonella enteritidis by means of bacteriophage**. *APMIS* 1950, **27**(4):625-640.
32. Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, Wehnes C, Van Kessel JS, Karns JS, Musser SM, Brown EW: **Phylogenetic Diversity of the Enteric Pathogen Salmonella enterica subsp. enterica Inferred from Genome-Wide Reference-Free SNP Characters**. *Genome Biology and Evolution* 2013, **5**(11):2109-2123.
33. Moreno Switt AI, Soyer Y, Warnick LD, Wiedmann M: **Emergence, Distribution, and Molecular and Phenotypic Characteristics of Salmonella enterica Serotype 4, 5, 12: i:–**. *Foodborne pathogens and disease* 2009, **6**(4):407-415.
34. De la Torre E, Zapata D, Tello M, Mejia W, Frias N, Pena FJG, Mateu EM, Torre E: **Several Salmonella enterica subsp. enterica serotype 4, 5, 12: i:− phage types isolated from swine samples originate from serotype Typhimurium DT U302**. *Journal of clinical microbiology* 2003, **41**(6):2395-2400.
35. Benassi F, Martínez VF, Eiguer T: **Salmonella: its incidence in waters of the Zaiman arroyo**. *Revista Argentina de microbiologia* 1982, **15**(3):169-175.
36. Franco A, Hendriksen RS, Lorenzetti S, Onorati R, Gentile G, Dell'Omo G, Aarestrup FM, Battisti A: **Characterization of Salmonella occurring at high prevalence in a population of the land iguana Conolophus subcristatus in Galápagos Islands, Ecuador**. *PLoS One* 2011, **6**(8):e23147.
37. Olsen SJ, Bishop R, Brenner FW, Roels TH, Bean N, Tauxe RV, Slutsker L: **The changing epidemiology of Salmonella: trends in serotypes isolated from humans in the United States, 1987–1997**. *Journal of Infectious Diseases* 2001, **183**(5):753-761.

38. Bertelloni F, Chemaly M, Cerri D, Gall FL, Ebani VV: **Salmonella infection in healthy pet reptiles: Bacteriological isolation and study of some pathogenic characters**. *Acta Microbiologica et Immunologica Hungarica* 2016, **63**(2):203-216.
39. Woodward DL, Khakhria R, Johnson WM: **Human salmonellosis associated with exotic pets**. *Journal of Clinical Microbiology* 1997, **35**(11):2786-2790.
40. Sadler-Reeves L, Aird H, Pinna E, Elviss N, Fox A, Kaye M, Jorgensen F, Lane C, Willis C, McLauchlin J: **The occurrence of Salmonella in raw and ready-to-eat bean sprouts and sprouted seeds on retail sale in England and Northern Ireland**. *Letters in applied microbiology* 2016, **62**(2):126-129.
41. Darling ACE, Mau B, Blattner FR, Perna NT: **Mauve: multiple alignment of conserved genomic sequence with rearrangements**. *Genome research* 2004, **14**(7):1394-1403.
42. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT: **The evolution of metabolic networks of E. coli**. *BMC systems biology* 2011, **5**(1):182.
43. Thiele I, Hyduke DR, Steeb B, Fankam G, Allen DK, Bazzani S, Charusanti P, Chen F-C, Fleming RM, Hsiung CA: **A community effort towards a knowledge-base and mathematical model of the human pathogen Salmonella Typhimurium LT2**. *BMC systems biology* 2011, **5**(1):8.
44. Becker SA, Palsson BØ: **Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation**. *BMC microbiology* 2005, **5**(1):8.
45. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, Yi KY, Rhee JH, Lee SY: **Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery**. *Molecular systems biology* 2011, **7**(1):460.
46. McCloskey D, Palsson BØ, Feist AM: **Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli**. *Molecular systems biology* 2013, **9**(1):661.
47. AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP, Maranas CD, Woodward MJ, Anjum MF: **Genome scale reconstruction of a salmonella metabolic model comparison of similarity and differences with a commensal escherichia coli strain**. *Journal of Biological Chemistry* 2009, **284**(43):29480-29488.
48. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S: **Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction**. *BMC systems biology* 2009, **3**(1):38.
49. Oberhardt MA, Palsson BØ, Papin JA: **Applications of genome-scale metabolic reconstructions**. *Molecular systems biology* 2009, **5**(1):320.
50. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M: **The RAST Server: rapid annotations using subsystems technology**. *BMC genomics* 2008, **9**(1):75.
51. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nature biotechnology* 2010, **28**(9):977-982.
52. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes**. *Nucleic acids research* 2005, **33**(17):5691-5702.
53. Latendresse M: **Efficiently gap-filling reaction networks**. *BMC bioinformatics* 2014, **15**(1):225.
54. Thiele I, Palsson BØ: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nature protocols* 2010, **5**(1):93-121.

55. Larocque M, Chénard T, Najmanovich R: **A curated C. difficile strain 630 metabolic network: prediction of essential targets and inhibitors**. *BMC systems biology* 2014, **8**(1):117.
56. **United States Code Title 7**. In.; 2011.
57. Rigby D, Cáceres D: **Organic farming and the sustainability of agricultural systems**. *Agricultural systems* 2001, **68**(1):21-40.
58. Chen Z, Jiang X: **Microbiological safety of chicken litter or chicken litter-based organic fertilizers: a review**. *Agriculture* 2014, **4**(1):1-29.
59. Jung K-S, Heu S-G, Roh E-J, Kim M-H, Gil H-J, Choi N-Y, Lee D-H, Lim J-A, Ryu J-G, Kim K-H: **Survival of Salmonella enterica and Listeria monocytogenes in Chicken and Pig Manure Compost**. *Korean Journal of Soil Science and Fertilizer* 2013, **46**(6):469-473.
60. Wilkinson K, Tee E, Tomkins R, Hepworth G, Premier R: **Effect of heating and aging of poultry litter on the persistence of enteric bacteria**. *Poultry science* 2011, **90**(1):10-18.
61. **Climate of Minnesota**. In., 5/22/2017 edn. Wikipedia: The Wikimedia Foundation, Inc.; 2017.
62. Larney FJ, Yanke LJ, Miller JJ, McAllister TA: **Fate of coliform bacteria in composted beef cattle feedlot manure**. *Journal of environmental quality* 2003, **32**(4):1508-1515.
63. Guard-Petter J: **The chicken, the egg and Salmonella enteritidis**. *Environmental microbiology* 2001, **3**(7):421-430.
64. Bortolussi R: **Listeriosis: a primer**. *Canadian Medical Association Journal* 2008, **179**(8):795-797.
65. Lund BM, O'Brien SJ: **The occurrence and prevention of foodborne disease in vulnerable people**. *Foodborne pathogens and disease* 2011, **8**(9):961-973.
66. **BAM: Salmonella** [https://www.fda.gov/food/foodscienceresearch/laboratorymethods/ucm070149.htm]
67. **Annual Weather Averages Near Minneapolis**. In. www.timeanddate.com.
68. Control CfD, Prevention: **Ongoing multistate outbreak of Escherichia coli serotype O157: H7 infections associated with consumption of fresh spinach--United States, September 2006**. *MMWR Morbidity and mortality weekly report* 2006, **55**(38):1045.
69. Hudson J, Mott S: **Growth of Listeria monocytogenes, Aeromonas hydrophila and Yersinia enterocolitica on cold-smoked salmon under refrigeration and mild temperature abuse**. *Food Microbiology* 1993, **10**(1):61-68.
70. Control CfD, Prevention: **Multistate outbreak of listeriosis associated with Jensen Farms cantaloupe--United States, August-September 2011**. *MMWR Morbidity and mortality weekly report* 2011, **60**(39):1357.
71. Kim J, Shepherd Jr MW, Jiang X: **Evaluating the effect of environmental factors on pathogen regrowth in compost extract**. *Microb Ecol* 2009, **58**(3):498-508.
72. **FSMA Final Rule on Produce Safety** [https://www.fda.gov/food/guidanceregulation/fsma/ucm334114.htm]
73. **Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens**. In. Leafy Greens Marketing Agreement; 2010: 25.
74. Nicholson FA, Groves SJ, Chambers BJ: **Pathogen survival during livestock manure storage and following land application**. *Bioresource technology* 2005, **96**(2):135-143.
75. Himathongkham S, Riemann H: **Destruction of Salmonella typhimurium, Escherichia coli O157: H7 and Listeria monocytogenes in chicken manure by drying and/or gassing with ammonia**. *FEMS microbiology letters* 1999, **171**(2):179-182.
76. Rohmer L, Hocquet D, Miller SI: **Are pathogenic bacteria just looking for food?**

Metabolism and microbial pathogenesis. *Trends Microbiol* 2011, **19**(7):341-348.
77. Muñoz-Elías EJ, McKinney JD: **Carbon metabolism of intracellular bacteria**. *Cellular microbiology* 2006, **8**(1):10-22.
78. Tazzyman SJ, Bonhoeffer S: **Why there are no essential genes on plasmids**. *Molecular biology and evolution* 2015, **32**(12):3079-3088.
79. Charusanti P, Chauhan S, McAteer K, Lerman JA, Hyduke DR, Motin VL, Ansong C, Adkins JN, Palsson BO: **An experimentally-supported genome-scale metabolic network reconstruction for Yersinia pestis CO92**. *BMC Systems Biology* 2011, **5**(1):163.
80. Liao YC, Huang TW, Chen FC, Charusanti P, Hong JSJ, Chang HY, Tsai SF, Palsson BO, Hsiung CA: **An Experimentally Validated Genome-Scale Metabolic Reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228**. *J Bacteriol* 2011, **193**(7):1710-1717.
81. Oberhardt MA, Puchałka J, Fryer KE, Martins dos Santos VAP, Papin JA: **Genome-Scale Metabolic Network Analysis of the Opportunistic Pathogen Pseudomonas aeruginosa PAO1**. *J Bacteriol* 2008, **190**(8):2790-2803.
82. Pinchuk GE, Hill EA, Geydebrekht OV, De Ingeniis J, Zhang X, Osterman A, Scott JH, Reed SB, Romine MF, Konopka AE *et al*: **Constraint-Based Model of Shewanella oneidensis MR-1 Metabolism: A Tool for Data Analysis and Hypothesis Generation**. *PLoS Comput Biol* 2010, **6**(6).
83. Baumler DJ, Ma B, Reed JL, Perna NT: **Inferring ancient metabolism using ancestral core metabolic models of enterobacteria**. *BMC Systems Biology* 2013, **7**(1):46.
84. Bussieck MR, Meeraus A: **General algebraic modeling system (GAMS)**. *Applied Optimization* 2004, **88**:137-158.
85. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M: **The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)**. *Nucleic acids research* 2014, **42**(D1):D206-D214.
86. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nat Biotech* 2010, **28**(9):977-982.
87. Sutton VR, Kiley PJ: **Techniques for studying the oxygen-sensitive transcription factor FNR from Escherichia coli**. *Methods in enzymology* 2003, **370**:300-312.
88. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110**. *Applied and environmental microbiology* 1994, **60**(10):3724-3731.
89. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT: **The evolution of metabolic networks of E. coli**. *BMC Systems Biology* 2011, **5**:182-182.
90. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW: **Genomic islands: tools of bacterial horizontal gene transfer and evolution**. *FEMS microbiology reviews* 2009, **33**(2):376-393.
91. Cusumano ZT, Caparon MG: **Citrulline protects Streptococcus pyogenes from acid stress using the arginine deiminase pathway and the F1Fo-ATPase**. *Journal of bacteriology* 2015, **197**(7):1288-1296.
92. Bochner BR: **Global phenotypic characterization of bacteria**. *FEMS microbiology reviews* 2009, **33**(1):191-205.
93. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: Rapid Annotations using Subsystems Technology**. *BMC Genomics* 2008, **9**:75.